



# A micromechanics-inspired constitutive model for shape-memory alloys that accounts for initiation and saturation of phase transformation



Alex Kelly<sup>a</sup>, Aaron P. Stebner<sup>b</sup>, Kaushik Bhattacharya<sup>a,\*</sup>

<sup>a</sup> Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA 91125, USA

<sup>b</sup> Mechanical Engineering, Colorado School of Mines, Golden, CO 80401, USA

## ARTICLE INFO

### Article history:

Received 29 May 2015

Received in revised form

2 February 2016

Accepted 17 February 2016

Available online 18 February 2016

### Keywords:

Shape-memory alloys

Constitutive model

Phase transformation

## ABSTRACT

A constitutive model to describe macroscopic elastic and transformation behaviors of polycrystalline shape-memory alloys is formulated using an internal variable thermodynamic framework. In a departure from prior phenomenological models, the proposed model treats initiation, growth kinetics, and saturation of transformation distinctly, consistent with physics revealed by recent multi-scale experiments and theoretical studies. Specifically, the proposed approach captures the macroscopic manifestations of three micromechanical facts, even though microstructures are not explicitly modeled: (1) Individual grains with favorable orientations and stresses for transformation are the first to nucleate martensite, and the local nucleation strain is relatively large. (2) Then, transformation interfaces propagate according to growth kinetics to traverse networks of grains, while previously formed martensite may reorient. (3) Ultimately, transformation saturates prior to 100% completion as some unfavorably-oriented grains do not transform; thus the total transformation strain of a polycrystal is modest relative to the initial, local nucleation strain. The proposed formulation also accounts for tension–compression asymmetry, processing anisotropy, and the distinction between stress-induced and temperature-induced transformations. Consequently, the model describes thermoelastic responses of shape-memory alloys subject to complex, multi-axial thermo-mechanical loadings. These abilities are demonstrated through detailed comparisons of simulations with experiments.

© 2016 Elsevier Ltd. All rights reserved.

## 1. Introduction

Shape-memory alloys (SMAs) exhibit unusual macroscopic phenomena including superelasticity, the shape-memory effect, and actuation. Through these behaviors, inelastic strains on the order of several percent are recovered because they are accommodated through diffusionless, reversible microstructure rearrangement instead of slip, glide, climb, and other irrecoverable, plastic mechanisms. These remarkable phenomena are enabled by solid-to-solid phase transformation between a high symmetry *austenite* phase that is stable at high temperatures, and a low symmetry *martensite* phase that is stable at low temperatures. The symmetry disparity between the phases allows multiple *variants* of the martensite phase – martensite

\* Corresponding author.

E-mail address: [bhatta@caltech.edu](mailto:bhatta@caltech.edu) (K. Bhattacharya).

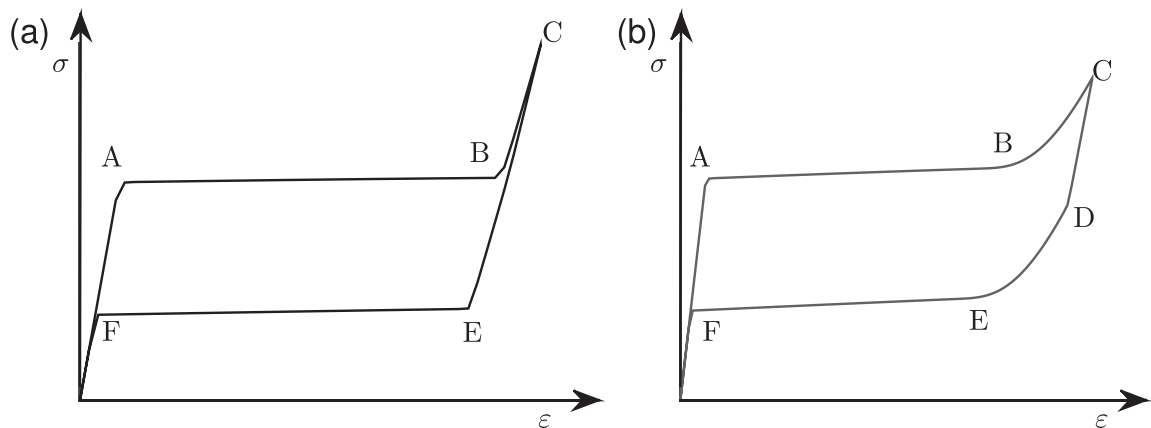
structures that are identical up to their orientation – to form from each austenite structure during transformation. The ability to form variants or patterns of variants manifests an infinite number of possibilities for macroscopic strains, or shape changes, during transformation of a polycrystal. As for a liquid–solid or liquid–gas transformation, a Clausius–Clapeyron relation governs the solid-state phase transformation of SMAs. Thus, the transformation may be induced by changing temperature or stress. The state of thermo-mechanical loading during transformation will also dictate the shape change. Furthermore, while in the martensitic state, one variant may switch to another – *i.e.*, *reorient* – in response to applied forces, also resulting in macroscopic shape changes.

Regardless of whether a given martensite structure was formed directly during transformation or via post-transformation deformation, upon reverse transformation, all variants of martensite revert back to the same initial austenite structure, resulting in full shape recovery, or “shape-memory.” *Superelasticity* describes a constant ambient temperature event in which elastic and transformation strains induced by mechanical loading are completely recovered upon unloading. Analogously, via the *shape-memory effect*, a SMA mechanically deformed, then unloaded below the critical transformation temperature, while in the martensite phase, returns to its original shape upon heating to austenite. *Actuation* is achieved by maintaining an applied load while heating and cooling through transformation; the SMA recovers transformation strain against the load thereby producing mechanical work.

The unusual properties of SMAs have led to their use in a variety of engineering applications ranging from implantable medical devices to actuators. We refer the reader to [Otsuka and Wayman \(1998\)](#) for a detailed introduction. The increasingly sophisticated applications of these materials calls for a model that is capable of faithfully describing complex phenomena, while being simple enough to be used in the engineering design process. This demand has motivated a number of models. Some of these, like the widely used models of [Auricchio et al. \(1997\)](#), [Auricchio and Petrini \(2004\)](#) and [Lexcellent et al. \(2002\)](#), are phenomenological and adapt frameworks from other subjects like plasticity. Others, like the models of [Qidwai and Lagoudas \(2000\)](#), [Panico and Brinson \(2007\)](#), [Sadjadpour and Bhattacharya \(2007b,a\)](#), and [Chemisky et al. \(2011\)](#) introduce internal variables to incorporate some microscopic information. Yet other models like those of [Zaki and Moumni \(2007\)](#) focus on thermomechanical interactions and cyclic loadings. We refer the reader to [Chemisky et al. \(2011\)](#) for a comprehensive survey of the literature. Despite significant advances, a phenomenological model that accurately describes transformation initiation, growth, and saturation during multiaxial proportional and non-proportional loading remains a work in progress, and motivates the proposed model.

We now consider superelasticity in some detail to describe the ideas of the current model. To understand the underlying mechanisms behind this phenomenon, we first consider an ideal *single crystal* at a temperature just above its transformation temperature. At this temperature, austenite is the stable phase and martensite the metastable phase. As the crystal is subjected to stress, it initially responds elastically as in loading prior to point A in [Fig. 1a](#). However, at some critical stress, the martensite is stabilized ([Burkart and Read, 1953](#); [James, 1986](#)) and the material begins to transform from austenite to martensite. This event results in a macroscopic inelastic yield event that appears similar to perfect plasticity (A). Once the material is fully martensite, the martensite now responds elastically on further loading to some maximum load below the stress required for detwinning and/or plastic flow (B to C). Upon unloading, the martensite remains the stabilized phase as long as the stress remains high and unloads elastically (C–E). However, at some critical stress, it becomes unstable and begins to transform to the austenite, and this reverse transformation gives rise to the lower plateau (E to F). Once the transformation is complete, the austenite unloads elastically.

Hysteresis results from energy dissipated during transformation. For superelasticity, it is often characterized by the differences between the critical stresses for forward and reverse transformation. Single crystal transformation responses differ with differing loading directions due to crystallographic anisotropy ([Miyazaki et al., 1984](#); [Shield, 1995](#)). Specifically, a



**Fig. 1.** A schematic representation of superelasticity in (a) single crystal and (a) polycrystalline specimens.

critically resolved stress criterion adequately describes the single crystal anisotropy of transformation (Miyazaki et al., 1984; Shield, 1995). Asymmetries with respect to loading mode (e.g., tension vs. compression) and the deformation mechanisms are broadly explained by considering the Gibbs energy, and more specifically the Clausius–Clapeyron relation between the critical stress for transformation, the strain of transformation, and temperature (Burkart and Read, 1953; James, 1986).

The situation is considerably more complicated in a *polycrystal*. The transformation strains, microstructures, and responses to the applied stress of the grains all depend on their orientation. Thus unique grains attempt to deform uniquely. However, they are not free to do so due to intergranular constraints. Therefore transformation is a complex composite process in a polycrystal; the subject of many recent experimental and theoretical investigations (see for example, Sitter et al., 2010; Stebner et al., 2013). Furthermore, the quantified characteristics of superelastic responses vary significantly with material, processing, and loading mode. Non-proportional loading provides further complexities (Thamburaja and Anand, 2002; McNaney et al., 2003; Richards et al., 2013). The models we previously mentioned each focus on certain aspects of these phenomena with varying levels of complexity and fidelity, but an engineering demand for a simple, unified framework that describes all of these observations still exists.

To understand how the present model aims to address this demand, we now consider a polycrystal completely in the austenite state, above its transformation temperature, subjected to increasing tensile stress. A characteristic polycrystal superelastic stress–strain curve in this circumstance is shown in Fig. 1b. The material initially loads elastically, then yields to transformation at the point marked A, macroscopically similar to the single crystal case. The stress in the polycrystal is largely uniform during elastic loading except as a result of the elastic anisotropy of the austenite. Thus, if we were to ignore the latter, the first grains that will transform are those most favorably oriented for transformation according to a resolved stress criterion, as observed in recent neutron diffraction experiments by Stebner et al. (2015). Using in-situ optical microscopy, Brinson et al. (2004) have also observed that the first appearance of the martensite occurs in isolated regions in well-oriented grains. This finding is also supported by the mesoscale observations of Daly et al. (2008) that deviations in linearity of the stress–strain curve occurs well before the formation of macroscopic transformed regions. Bhattacharya and Schlömerkemper (2010) have recently proved in an idealized setting with uniform modulus that the transformation begins in isolated grains, while Richards et al. have made the same observation using a full-field micromechanical model (Richards et al., 2013). Finally, Sittner and Novak (2000) have noted that the constant stress Sachs or Ruess bound accurately describes the initiation of the transformation in polycrystals. In summary, transformation *initiation* is governed by well-oriented grains.

Transformation growth kinetics, however, are not. As the best oriented grains begin to transform, whole grains cannot due to the constraints of the neighbors and the transformation proceeds in isolated regions (Richards et al., 2013). Thus, the internal stress distribution amongst grains becomes extremely heterogeneous as stress is redistributed from ideally oriented grains where transformation has nucleated to slightly misoriented grains, which subsequently nucleate (Richards et al., 2013; Paranjape and Anderson, 2014). These events cause a rapid progression of the transformation within the polycrystal. This progression is further accelerated by the fact that once transformation has nucleated in a grain, little additional driving force is needed for martensite growth. Additionally, stress redistribution may change the stress orientation within transformed grains, and these events will cause martensite variants other than those initially formed to be preferred (Stebner et al., 2015). The activity of these mechanisms in tandem gives rise to the stress–strain plateau AB.

Macroscopically, single and polycrystal superelasticity have appeared very similar up to point B. However, in loading beyond the plateau (BC), polycrystals exhibit pronounced hardening (Fig. 1b) where most single crystals do not (Fig. 1a). In polycrystals loaded to point B (Fig. 1b), transformation is incomplete or non-existent in some of the most poorly oriented grains. Further transformation in these grains requires increased stress, while grains that have completed transformation begin to elastically load again. Twinning events are also incited that reorient martensite to a greater extent than in the plateau region (Stebner et al., 2015). All of these mechanisms in combination give rise to the observed hardening BC. As a result, the transformation of the polycrystal is quite heterogeneous (Brinson et al., 2004; Daly et al., 2008; Richards et al., 2013; Stebner et al., 2015; Paranjape and Anderson, 2014), as is the state of internal stress (Pelton et al., 2015). As transformation and reorientation proceed, the poorly oriented grains exhibit smaller transformation strain in the direction of loading, and thus they begin to saturate and lock together eventually leading to a network of fully transformed grains surrounded by partially transformed and untransformed grains at C (Brinson et al., 2004). Thus, the *saturation* of the transformation is governed by the poorly oriented grains. This phenomenon is also the reason why the constant strain Taylor or Voigt bound gives a good description of the macroscopic transformation strains (Bhattacharya and Kohn, 1997; Shu and Bhattacharya, 1998; Bhattacharya and Suquet, 2005).

Forward loading is continued until some maximum load is reached at point C; further loading evokes plasticity, which is beyond the scope of the current model. Upon unloading, the material initially unloads elastically. However, at some point, marked D, the material begins to recover its inelastic strain, initially with some softening and then (point E) perfectly. Upon the recovery of all inelastic strain, at point F, the material unloads elastically.

An interesting consequence of the fact that the initiation and saturation of the martensitic transformation in polycrystals are governed by two separate mechanisms is that the critically resolved shear stress criterion or the stress–strain aspect of the Clausius–Clapeyron relation fails in a polycrystal after the very first initiation(s) of transformation (Daly et al., 2007). The *key heuristic and the point of departure* for the current work is the explicit recognition that the mechanics of initiation and saturation of the martensitic transformation in polycrystals are two essentially different processes. Therefore, we differ-

entiate between two separate kinematic quantities, the nominal transformation strain in the martensite and the effective transformation strain of the polycrystal. The former heuristically represents the macroscopic transformation strain one would see if the grains were allowed to transform independently. This quantity governs the initiation of transformation. The latter heuristically represents the transformation strain one actually sees limited by intergranular constraints. This quantity governs the saturation of the transformation. We will see that the reorientation of both stress-induced and thermally stabilized martensite arises naturally due to competition between these mechanisms. While, for simplicity, the introduction above focused on superelasticity and stress-induced martensite, the model we present is thermomechanical and also capable of describing thermal transformations and the resulting shape-memory effect and actuation phenomena.

The present model builds on the framework developed by [Sadjadpour and Bhattacharya \(2007b,a\)](#). Their work, like other internal variable models in the literature, did not distinguish between initiation and saturation mechanisms. The current work does, and also differs by adopting a new formulation for transformation surfaces that follows from a formula introduced by [Cazacu et al. \(2006\)](#) to describe anisotropic and asymmetric plastic yield surfaces; this change improves the ability to capture transformation asymmetry and processing-induced material anisotropy of SMAs. The present model has similarities to those of [Panico and Brinson \(2007\)](#) and [Chemisky et al. \(2011\)](#). Those models make a distinction between self-accommodated and oriented martensite, hence they capture some of the features describe above. However, they do not completely describe the difference between initiation and saturation.

The kinetics of transformation and reorientation in the model we present are rate-independent. Relative to the slow loading rates of typical engineering applications ( $10^{-2} \text{ s}^{-1}$  and below), these kinetics are largely rate independent ([Abeyaratne et al., 1996](#)). Macroscopic rate effects are often seen, but these are often the consequence of the release of latent heat, which is not inherent to the kinetics. In the present model, apparent rate effects due to thermal energy enter naturally via the energy balance. On occasion, rate effects are also introduced due to grips, work hardened alloys, etc., but these are small and are the result of particular circumstances that we do not address in this work. We note, however, that inherent rate effects are important in dynamic situations, and the model can easily be modified to account for these if necessary (for example as in [Sadjadpour and Bhattacharya, 2007b,a](#)).

Finally, some (highly worked) specimens of nickel–titanium based SMAs (a.k.a. “Nitinol”) display R-phase, an intermediate phase that forms in between the austenite and martensite phases during thermal cycling ([Otsuka and Wayman, 1998](#), [Duerig and Bhattacharya, 2015](#)). This phase has a small transformation strain and only appears in engineering applications with very particular processing and thermomechanical loading conditions. We ignore the R-phase in this model.

We proceed to introduce the model in [Section 2](#), analytically demonstrate some features of the model in simplified settings in [Section 3](#), discuss our numerical implementation in [Section 4](#), and conduct a parameter study while also demonstrating the model relative to experimental data in [Section 5](#). We conclude with a discussion of desirable future work with the framework in [Section 6](#).

## 2. Continuum model

### 2.1. Kinematics

Consider a macroscopic polycrystalline specimen of a shape-memory alloy, so that each material point corresponds to a representative volume of material with numerous austenite grains, with possible fine-scale martensite microstructure within each grain. We denote the macroscopic strain as  $\varepsilon$  and the (absolute) temperature as  $\theta$ . As in [Sadjadpour and Bhattacharya \(2007b\)](#), we introduce two additional internal variables that describe the state of transformation. The first is the *volume fraction* of the martensite  $\lambda$ . By definition, this variable satisfies the constraint

$$0 \leq \lambda \leq 1. \quad (1)$$

The second is the *nominal transformation strain* of the martensite  $\varepsilon_m$ , which we define as the average transformation strain of every region of martensite in every grain in the representative volume. We will assume that

$$\varepsilon_m \in \mathcal{G}_i \quad (2)$$

where  $\mathcal{G}_i$  describes the set of all possible nominal transformation strains that may occur. This set could be computed by taking averages of local martensite strains over all variants in the representative volume, however we prescribe this phenomenologically in the next section. In subsequent sections of the model formulation,  $\varepsilon_m$  will dictate the driving force on transformation and the set  $\mathcal{G}_i$  will define initiation events. The overall or *effective transformation strain* is obtained by multiplying the nominal transformation strain with the volume fraction,  $\lambda \varepsilon_m$ . This overall transformation strain drives macroscopic behaviors governed by inter-granular compatibility; thus  $\lambda \varepsilon_m$  will play a critical role in saturation. We define a set  $\mathcal{G}_s$  that describes the set of effective transformation strains that may be achieved without hardening/softening influence from poorly oriented grains in forward/reverse transformation events. The effective transformation strain may take on values outside of this set with some energy penalty, as mathematically prescribed in the next section.

## 2.2. Energy

The Helmholtz free energy density of the system is postulated to be of the form

$$W(\varepsilon, \varepsilon_m, \lambda, \theta) = \frac{1}{2}(\varepsilon - \lambda \varepsilon_m) : C(\lambda) : (\varepsilon - \lambda \varepsilon_m) + \lambda L \frac{(\theta - \theta_c)}{\theta_c} - c_p(\lambda) \theta \ln \left( \frac{\theta}{\theta_c} \right) + \lambda G_i(\varepsilon_m) + G_s(\lambda \varepsilon_m) + G_\lambda(\lambda). \quad (3)$$

The first term is the elastic energy density, which depends quadratically on the elastic strain, i.e., the difference between the total strain and the effective transformation strain. The second term represents the excess chemical energy density of martensite which depends on temperature. Note that we have chosen to restrict the model to a moderate range of temperatures close to the thermodynamic transformation temperature  $\theta_c$  where this relation may be taken as linear, scaling with the latent heat of transformation,  $L$ , relative to the difference between the absolute temperature and the thermodynamic transformation temperature. The third term is the contribution of heat capacity  $c_p$ . Each phase of the material may exhibit a unique heat capacity as well as a unique tensor of bulk elastic constants,  $C$ , thus we adopt a rule of mixtures approach in defining the composite responses of the two phases at a material point<sup>1</sup>

$$C(\lambda) = \lambda C^M + (1 - \lambda) C^A, \quad (4)$$

$$c_p(\lambda) = \lambda c_p^M + (1 - \lambda) c_p^A, \quad (5)$$

where the superscripts  $A$  and  $M$  denote austenite and martensite phase properties, respectively.

The last three terms in the energy  $G_i$ ,  $G_s$ , and  $G_\lambda$  describe the energy densities for initiation, saturation, and growth of transformation, respectively. Specifically,  $G_\lambda$  reflects internal stored energy related to incompatibility between variants and across grains; i.e., the barrier to transformation growth. Thus, in a manner analogous to hardening in plasticity, we postulate that<sup>2</sup>

$$G_\lambda(\xi) = \begin{cases} g_\lambda(\xi) = \frac{f_{\lambda\lambda}}{n_{\lambda\lambda} + 1} \xi^{n_{\lambda\lambda} + 1} & \xi \in [0, 1] \\ +\infty & \text{else} \end{cases} \quad (6)$$

where  $f_{\lambda\lambda} > 0$ ,  $n_{\lambda\lambda} > 0$  are constitutive constants. The first line accounts for the fact that the amount of stored energy in the system can increase with increasing volume fraction due to the development of internal stresses, thus for choices of  $f_{\lambda\lambda} \neq 0$ ,  $\lambda$  experiences some hardening within the set  $[0, 1]$ . The second line enforces the hard constraint (1).

The functions  $G_\alpha$  for  $\alpha = i, s$  are defined over symmetric matrices. To prescribe the sets of nominal and effective transformation strains, we introduce functions of the form  $g_\alpha$  following Cazacu et al. (2006):

$$g_\alpha(\xi) = \left( \sum_{n=1}^3 \left( |\mu_n(C_\alpha : \xi)| - a_{1\alpha} \mu_n(C_\alpha : \xi) \right)^{a_{2\alpha}} \right) - b_\alpha^{a_{2\alpha}} \quad (7)$$

where  $\xi$  is any symmetric matrix,  $\mu_n(\xi)$ ,  $n = 1, 2, 3$  denote the principal values of the matrix argument  $\xi$ ,  $C_\alpha$  are 4th order tensors satisfying minor and major symmetries and reflecting the anisotropy of the material (choosing  $C_\alpha = I$  where  $I$  is the 4th order identity tensor results in isotropy) and  $a_{1\alpha}$ ,  $a_{2\alpha}$ ,  $b_\alpha$  are constitutive constants to be prescribed. We assume that  $-1 < a_{1\alpha} < 1$ ,  $a_{2\alpha} > 1$ ,  $b_\alpha > 0$  to ensure the convexity and coercivity of  $g_\alpha$ .

We postulate

$$G_i(\xi) = \begin{cases} 0 & \xi \in \mathcal{G}_i \\ +\infty & \text{else.} \end{cases} \quad (8)$$

where the set of allowable initiation strains is taken as<sup>3</sup>

$$\mathcal{G}_i = \{\xi : \text{tr}(\xi) = 0, g_i(\xi) \leq 0\}. \quad (9)$$

Eq. (8) enforces the constraint (2). The martensite strain may evolve freely with no energy cost within the set  $\mathcal{G}_i$ , but evolution outside of this set is not allowed. We can also give physical meaning to the coefficients in (7). The coefficient  $b_i$  denotes some radius of the set  $\mathcal{G}_i$  while the constants  $a_{1i}$  and  $a_{2i}$  control asymmetry in the same way the coefficients  $k$  and  $a$  do in the formulation of Cazacu et al. (see Fig. 4 of Cazacu et al., 2006).

We note that the expressions (7) are different from the polynomials of the invariants  $J_2$  and  $J_3$  used in Sadjadpour and Bhattacharya (2007a). We found that these formulae restricted the amount of asymmetry that one could prescribe before

<sup>1</sup> We assume that the bulk moduli of the phases (under hydrostatic stress) are equal consistent with experimental observations of SMAs (Patel and Cohen, 1953).

<sup>2</sup> We introduce a notation where  $\xi$  is generically used to represent the argument of the functions; i.e., it is a placeholder for the internal variables of the model defined in (3).

<sup>3</sup> We clarify that the colon in  $\xi$ :  $\text{tr}(\xi)$  reads “ $\xi$  such that the trace of  $\xi$ ,” where in (7) the colon denotes the contraction of the fourth and second order tensors. Despite the ambiguity, both uses of the colon are obvious in their context.

the sets of allowable strains became non-convex, especially in anisotropic media. This limitation proved unsatisfactory for highly textured SMA polycrystals. In contrast, (7) enabled greater asymmetries of convex sets, especially for anisotropic media (Cazacu et al., 2006).

Recall that for saturation, the effect of incompatibility between variants and grains means that continued transformation strain requires internal stresses and additional stored energy. Thus, unlike the previous hard constraints on the allowable values of volume fraction and martensite strains, we postulate a soft constraint for saturation such that kinematic hardening occurs once the allowable set of effective transformation saturation strains is exceeded:

$$G_s(\xi) = \begin{cases} 0 & g_s(\xi) \leq 0, \text{ tr}(\xi) = 0 \\ k_s \left( \ln \left( \frac{g_s(\xi)}{b_s^{a_{2s}}} \right) \right)^{m_s} & g_s(\xi) > 0, \text{ tr}(\xi) = 0 \\ +\infty & \text{tr}(\xi) \neq 0 \end{cases} \quad (10)$$

where  $k_s$  and  $m_s$  are the saturation hardening coefficient and exponent, respectively. The function  $G_s=0$  on the set

$$\mathcal{G}_s = \{\xi: \text{tr}(\xi) = 0, g_s(\xi) \leq 0\} \subset \mathcal{G}_i \quad (11)$$

where the final inequality is enforced by the choice of constants  $b_s$ ,  $a_{1s}$ , and  $a_{2s}$ , which have analogous interpretation as their initiation counterparts.

Finally, we assume the usual Fourier law for heat conduction, which is a linear relation between heat flux  $q$  and temperature gradient according to the thermal conductivity,  $K$

$$q = -K \nabla \theta. \quad (12)$$

### 2.3. Balance laws, driving forces and kinetics

We postulate the usual balance laws for equilibrium and energy as well as the second law of thermodynamics (Clausius–Duhem inequality)

$$\text{div } \sigma + b = 0, \quad \sigma = \sigma^T, \quad (13)$$

$$\dot{\epsilon} = \sigma: \dot{\epsilon} - \nabla \cdot q + r, \quad (14)$$

$$\dot{\eta} \geq \nabla \cdot \left( -\frac{q}{\theta} \right) + \frac{r}{\theta} \quad (15)$$

where  $\sigma$  is the stress,  $b$  is the body force,  $\epsilon$  is the internal energy density,  $\eta$  is the entropy density,  $\theta$  is the absolute temperature,  $q$  is the heat flux and  $r$  is the radiative heating.

With the free energy density specified in (3), we can use the balance laws and arguments following Coleman and Noll (1963) to obtain constitutive relations for the stress ( $\sigma$ ) and entropy ( $\eta$ ):

$$\sigma = \frac{\partial W}{\partial \epsilon} = C(\lambda): (\epsilon - \lambda \epsilon_m), \quad (16)$$

$$\eta = -\frac{\partial W}{\partial \theta} = -\lambda \frac{L}{\theta_c} + c_p \left( 1 + \ln \left( \frac{\theta}{\theta_0} \right) \right). \quad (17)$$

We also obtain the following driving forces as the thermodynamic conjugates to the internal variables  $\lambda$  and  $\epsilon_m$ . This is a little tricky since the function  $G_\alpha$  are not smooth. However, it is first instructive to assume that the functions  $G_{i,s,\lambda}$  are smooth. In such situations,

$$d_\lambda = -\frac{\partial W}{\partial \lambda} = -\frac{1}{2}(\epsilon - \lambda \epsilon_m): \frac{\partial C}{\partial \lambda}: (\epsilon - \lambda \epsilon_m) + \frac{dc_p}{d\lambda} \theta \ln \left( \frac{\theta}{\theta_c} \right) + \sigma: \epsilon_m - L \frac{(\theta - \theta_c)}{\theta_c} - G_i(\epsilon_m) - \epsilon_m: \frac{\partial G_s}{\partial \xi} \Big|_{\lambda \epsilon_m} - \frac{\partial G_\lambda}{\partial \xi} \Big|_\lambda, \quad (18)$$

$$d_{\epsilon_m} = -\frac{\partial W}{\partial \epsilon_m} = \lambda \sigma - \lambda \frac{\partial G_i}{\partial \xi} \Big|_{\epsilon_m} - \lambda \frac{\partial G_s}{\partial \xi} \Big|_{\lambda \epsilon_m}. \quad (19)$$

In the non-smooth situation, we interpret these equations using usual methods of convex analysis (Ekeland and Témam, 1999) as follows

$$-d_\lambda - \frac{1}{2}(\epsilon - \lambda \epsilon_m): \frac{\partial C}{\partial \lambda}: (\epsilon - \lambda \epsilon_m) + \frac{dc_p}{d\lambda} \theta \ln \left( \frac{\theta}{\theta_c} \right) + \sigma: \epsilon_m - L \frac{(\theta - \theta_c)}{\theta_c} - G_i(\epsilon_m) \in \epsilon_m: \partial G_s(\lambda \epsilon_m) + \partial G_\lambda(\lambda), \quad (20)$$

$$-d_{\epsilon_m} + \lambda \sigma^{\text{dev}} \in \lambda \partial G_i(\epsilon_m) + \lambda \partial G_s(\lambda \epsilon_m) \quad (21)$$



where  $\sigma^{dev} = \sigma - \frac{\text{tr } \sigma}{3} I$  and  $\partial G_\alpha$  denotes the sub-differential of the  $G_\alpha$ .

To close the system, we relate the driving forces to the relevant rates using kinetic relations. In the rate-dependent smooth setting, we postulate

$$\dot{\lambda} = K_\lambda(d_\lambda), \quad \dot{\varepsilon}_m = K_m(d_{\varepsilon_m}) \quad (22)$$

for given functions  $K_m$  and  $K_\lambda$ . However, in this paper, we are interested in the rate-independent case where  $\lambda$  and  $\varepsilon_m$  evolve freely if and only if the corresponding driving forces reach a critical value. Therefore, the functions  $K_\lambda$  and  $K_m$  are neither smooth or invertible. So, we specify them in the form of Kuhn–Tucker conditions. The rate of change of the internal variables are constitutive functions of the driving forces such that transformation evolves according to:

$$\dot{\lambda} = 0 \text{ for } |d_\lambda| < d_\lambda^c, \quad d_\lambda \dot{\lambda} \geq 0, \quad |d_\lambda| \leq d_\lambda^c \text{ if } 0 \leq \lambda \leq 1 \quad (23)$$

and martensite strain:

$$\dot{\varepsilon}_m = 0 \text{ for } \left| \frac{1}{\lambda} d_m \right| < d_m^c, \quad \frac{1}{\lambda} d_m : \dot{\varepsilon}_m \geq 0, \quad \left| \frac{1}{\lambda} d_m \right| \leq d_m^c \text{ if } \varepsilon_m \in \mathcal{G}_i; \quad (24)$$

$$\text{Reset } \varepsilon_m = 0 \text{ when } \lambda = 0, \quad \theta \geq A_f \quad (25)$$

where  $d_\lambda^c$  is the critical driving force for transformation and  $d_m^c$  is the critical driving force for martensite variant reorientation. We may rewrite these conditions by introducing a dissipation potential

$$D(\dot{\lambda}, \dot{\varepsilon}_m) = d_\lambda^c |\dot{\lambda}| + \lambda d_m^c |\dot{\varepsilon}_m| = D_\lambda(\dot{\lambda}) + D_m(\dot{\varepsilon}_m). \quad (26)$$

Now kinetic equations may be rewritten as

$$d_\lambda \in \partial D_\lambda, \quad d_{\varepsilon_m} \in \partial D_m \quad (27)$$

subject to the reset condition (25).

Combining (27) with (20) and (21), we can write the evolution equations for  $\lambda$ ,  $\varepsilon_m$  as

$$0 \in \frac{1}{2}(\varepsilon - \lambda \varepsilon_m) : \frac{\partial C}{\partial \lambda} : (\varepsilon - \lambda \varepsilon_m) + \frac{dc_p}{d\lambda} \theta \ln \left( \frac{\theta}{\theta_c} \right) - \sigma : \varepsilon_m + L \frac{(\theta - \theta_c)}{\theta_c} + G_i(\varepsilon_m) + \varepsilon_m : \partial G_s(\lambda \varepsilon_m) + \partial G_\lambda(\lambda) + \partial D_m(\varepsilon_m), \quad (28)$$

$$0 \in -\lambda \sigma^{dev} + \lambda \partial G_i(\varepsilon_m) + \lambda \partial G_s(\lambda \varepsilon_m) + \partial D_\lambda(\dot{\lambda}) \quad (29)$$

subject to the reset condition (25).

We will further discuss the interpretation of these conditions through specific analytic examples of shape memory behaviors in Section 3, and their numerical implementation in Section 4. For now, we note that these equations are consistent with the variational structure  $0 \in \partial W + \partial D$  (Mielke, 2005; Yang et al., 2006).

Finally, since  $\varepsilon = W + \theta \eta$  we can use (3), (12), and (17) to rewrite the energy balance (14) as

$$c_p \dot{\theta} = \nabla \cdot K \nabla \theta + r + \frac{\theta L}{\theta_c} \dot{\lambda} + d_\lambda \dot{\lambda} + d_{\varepsilon_m} : \dot{\varepsilon}_m. \quad (30)$$

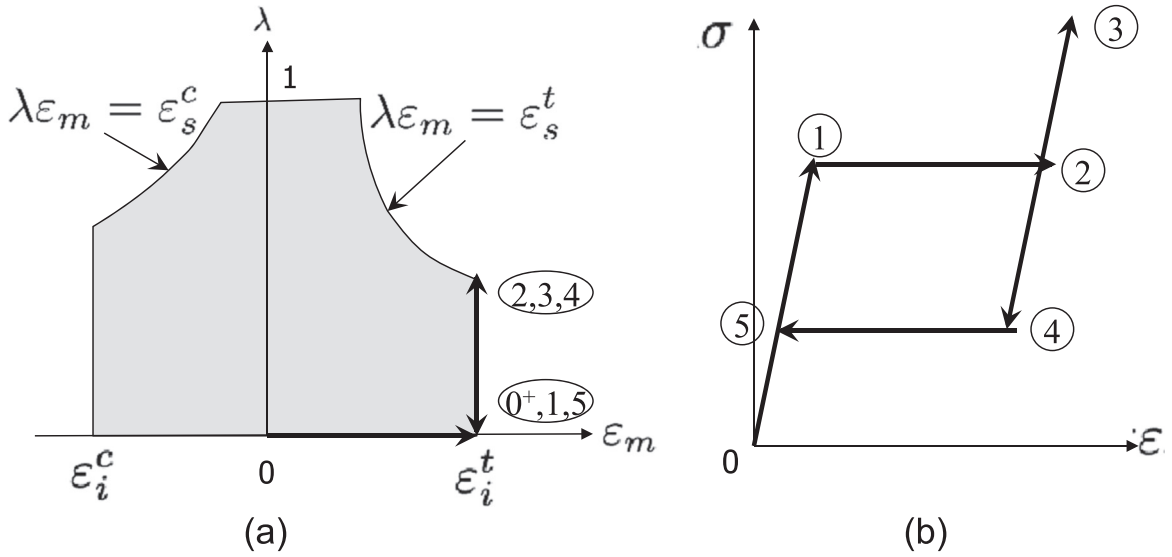
In summary, the model specifies that the behavior of SMAs is governed by the equilibrium equation (13), the heat equation (30) and the evolution equations (28) and (29).

### 3. Some features of the model

We now demonstrate some of the relationships between the parameters and mechanisms of the model through analytic calculations. To enable these analyses, we make a number of assumptions: we assume that the elastic moduli of the two phases  $C^A = C^M = C$  are equal and isotropic and the specific heat of the two phases  $c_p^A = c_p^M = c_p$  are equal. We also assume that the exponent  $m_s$  in the definition of  $G_s$  is infinite so that it becomes the hard constraint

$$G_s(\xi) = \begin{cases} 0 & \xi \in \mathcal{G}_i = \{\xi : \text{tr}(\xi) = 0, \quad g_s(\xi) \leq 0\}. \\ +\infty & \text{else.} \end{cases} \quad (31)$$

We also ignore the critical driving force for martensite reorientation ( $d_m^c = 0$ ). Finally, except for actuation in Section 3.1.3, we assume that  $f_{\lambda\lambda} = 0$  in (6) so that  $\lambda$  may evolve freely within the interval  $[0, 1]$ .



**Fig. 2.** The kinematic variables and stress-induced martensite in one dimension. (a) The volume fraction and nominal transformation strain are constrained to lie in the shaded region. (b) The stress–strain curve.

### 3.1. One dimension

It is instructive to specialize the model to one dimension. We can derive this one-dimensional model from the three dimensional formulation by assuming uniaxial loading; we reserve this formal derivation for [Section 3.3](#). For now, we simply assume

$$W(\epsilon, \epsilon_m, \lambda, \theta) = \frac{1}{2} E |\epsilon - \lambda \epsilon_m|^2 + \lambda \omega(\theta) - c_p \theta \ln\left(\frac{\theta}{\theta_0}\right) + \lambda G_i(\epsilon_m) + G_s(\lambda \epsilon_m) + G(\lambda), \quad (32)$$

where

$$\omega(\theta) = L \frac{(\theta - \theta_c)}{\theta_c} \quad (33)$$

and the sets  $\mathcal{G}_{i,s}$  reduce to the scalar intervals

$$\mathcal{G}_{i,s} = [\epsilon_{i,s}^c, \epsilon_{i,s}^t], \quad \epsilon_i^c \leq \epsilon_s^c \leq \epsilon_s^t \leq \epsilon_i^t. \quad (34)$$

The constraints that these choices impose on the kinematic variables  $\lambda$  and  $\epsilon_m$  are shown in [Fig. 2\(a\)](#).

#### 3.1.1. Superelasticity

Consider an isothermal strain-controlled experiment where the temperature is held constant at a value significantly higher than the transformation temperature so that  $\omega(\theta) > d_\lambda^c > 0$ . Consider the specimen at rest at zero stress so that it is in the austenite state with  $\lambda = 0$  and  $\epsilon_m = 0$ . Now subject the specimen to a monotonically increasing overall tensile strain  $\epsilon(t)$ . For very small times, the stress is given by  $\sigma(t) = E\epsilon(t)$ . Since this stress is positive, the nominal transformation strain  $\epsilon_m$  takes its maximal tensile value  $\epsilon_i^t$ . This is marked as the point  $0^+$  in [Fig. 2a](#). Following (18), the driving force for transformation in one-dimension is given by  $d_\lambda = E\epsilon\epsilon_i^t - \omega$ , and is initially negative. However, since  $\lambda$  is already zero, it can not decrease any further. Thus the volume-fraction remains zero and the stress increases proportional to the imposed strain according to  $E$ , and we begin to traverse from  $0^+/0$  toward the points marked 1 in [Fig. 2a/b](#).

As the applied strain increases, so do the stress and the driving force  $d_\lambda$  till the latter becomes positive and eventually reaches the value  $d_\lambda^c$  (i.e.,  $E\epsilon\epsilon_i^t - \omega = d_\lambda^c$ ) at the points marked 1 in the figure. At this point, the strain and stress are given as

$$\epsilon_{MS}^t = \frac{d_\lambda^c + \omega}{E\epsilon_i^t}, \quad \sigma_{MS}^t = \frac{d_\lambda^c + \omega}{\epsilon_i^t}. \quad (35)$$

Now transformation begins and proceeds in such a manner to keep  $d_\lambda$  constant and consequently the stress  $\sigma$  constant so that the material traverses from the points marked 1 towards 2 with increasing applied strain. As  $\lambda$  increases, the overall transformation strain  $\lambda\epsilon_m$  increases and eventually saturates the constraint  $\lambda\epsilon_m \leq \epsilon_s^t$  at the value



$$\lambda^t = \frac{\varepsilon_s^t}{\varepsilon_i^t}. \quad (36)$$

This is indicated by the points 2 in the figure, where

$$\varepsilon_{MF}^t = \frac{d_\lambda^c + \omega}{E\varepsilon_i^t} + \varepsilon_s^t, \quad \sigma_{MF}^t = \frac{d_\lambda^c + \omega}{\varepsilon_i^t}. \quad (37)$$

The transformation is now saturated, and further loading does not lead to any further transformation. Thus, the stress increases in an affine manner with applied strain and the material traverses towards the points 3 in the figure.

The specimen is then unloaded by monotonically decreasing the applied tensile strain. The transformation is initially constant, with no change, and the material unloads elastically, traversing towards the points 4 in the figure. The stress and the driving force decrease monotonically, till the latter becomes negative and eventually reaches the value  $-d_\lambda^c$  at points 4. Here,

$$\varepsilon_{AS}^t = \frac{-d_\lambda^c + \omega}{E\varepsilon_i^t} + \varepsilon_s^t, \quad \sigma_{AS}^t = \frac{-d_\lambda^c + \omega}{\varepsilon_i^t}. \quad (38)$$

Reverse transformation now begins at points 4 and  $\lambda$  decreases as unloading proceeds. The driving force and stress remain constant from points 4 to 5. The reverse transformation is complete at points 5 when  $\lambda = 0$ , and

$$\varepsilon_{AF}^t = \frac{-d_\lambda^c + \omega}{E\varepsilon_i^t}, \quad \sigma_{AF}^t = \frac{-d_\lambda^c + \omega}{\varepsilon_i^t}. \quad (39)$$

The material subsequently unloads elastically and the material returns to the origins.

We have an analogous situation in a compressive loading cycle with the analogous quantities obtained by replacing the superscripts  $t$  with  $c$  in the formulae (35)–(39) above.

A series of comments are in order. First note that the material parameters  $\varepsilon_i^{t,c}$  associated with the set of nominal transformation strains determine the initiation of transformation (c.f. (35)) while the material parameters  $\varepsilon_s^{t,c}$  associated with the set of effective transformation strains determine the extent of transformation (c.f. (37)). This justifies the subscript  $i$  for initiation for the former and the subscript  $s$  for saturation for the latter. More importantly, the parameters determining initiation and saturation are independently prescribed in this model.

Second and related, we can infer the values of transformation strain, the stress hysteresis and mean-value of stress to be

$$\begin{aligned} \varepsilon_{trans}^{t,c} &:= \frac{1}{2}(\varepsilon_{MF}^{t,c} - \varepsilon_{MS}^{t,c} + \varepsilon_{AS}^{t,c} - \varepsilon_{AF}^{t,c}) = \varepsilon_s^{t,c}, \\ \sigma_{hyst}^{t,c} &:= \frac{1}{2}(\sigma_{MS}^{t,c} + \sigma_{MF}^{t,c} - \sigma_{AF}^{t,c} - \sigma_{AS}^{t,c}) = \frac{2d_\lambda^c}{\varepsilon_i^{t,c}}, \\ \bar{\sigma}^{t,c} &:= \frac{1}{4}(\sigma_{MS}^{t,c} + \sigma_{MF}^{t,c} + \sigma_{AF}^{t,c} + \sigma_{AS}^{t,c}) = \frac{\omega}{\varepsilon_i^{t,c}}. \end{aligned} \quad (40)$$

Note that each of them can be independently constitutively prescribed – the transformation strains are determined by the parameters  $\varepsilon_s^{t,c}$ , the hysteresis by the parameter  $d_\lambda^c$  and the mean-value of the stress by  $\varepsilon_i^{t,c}$ , and there is no universal relation amongst them.

Third, tension and compression are different due to the difference in parameters with superscript  $t$  and superscript  $c$ . This tension-compression asymmetry is consistent with various observations going back to [Burkart and Read \(1953\)](#).

Fourth, note that the transformation from austenite to martensite does not go to completion, but saturates at a maximum volume fraction of  $\lambda^{t,c} = \varepsilon_i^{t,c}/\varepsilon_s^{t,c}$ . This is consistent with surface observations of [Brinson et al. \(2004\)](#), which have been recently affirmed to be a bulk phenomena by [Stebner et al. \(2015\)](#).

Fifth, the value of stress at which the transformation begins and completes or the reverse transformation begins and completes depend on temperature through  $\omega$ . If  $\omega$  depends linearly in temperature as in (33), then these stresses depend linearly on temperature consistent with the Clausius–Clapeyron relation ([Otsuka and Wayman, 1998](#)). Further, we can invert these relations to obtain the values of temperature where the transformation begins at zero stress:

$$M_s = M_f = \theta_c - \frac{d_\lambda^c}{L}, \quad A_s = A_f = \theta_c + \frac{d_\lambda^c}{L}. \quad (41)$$

Sixth, we see above that  $\sigma_{MS}^{t,c} = \sigma_{MF}^{t,c}$ ,  $\sigma_{AS}^{t,c} = \sigma_{AF}^{t,c}$  and  $M_s = M_f$ . These are all manifestations of the aforementioned assumptions. There is no reorientation of martensite because we have assumed one dimension. We shall see in [Section 3.2](#) that reorientation of martensite can give rise to a situation where  $\sigma_{MS}^{t,c} < \sigma_{MF}^{t,c}$ ,  $\sigma_{AS}^{t,c} > \sigma_{AF}^{t,c}$ . Here, we have also assumed hard constraints on the sets (34) and that  $f_{\lambda\lambda} = 0$ . The more general assumption or soft constraint (10) on  $G_s$  allows the transformation to continue to evolve to higher volume fractions with increasing stress so that  $\sigma_{MS}^{t,c} < \sigma_{MF}^{t,c}$ ,  $\sigma_{AS}^{t,c} > \sigma_{AF}^{t,c}$ . Setting  $f_{\lambda\lambda} > 0$  violates all three equalities (40). We will study these other choices extensively in [Section 5](#). We also assume rate-independent kinetics (23) and (24). Rate-dependent kinetics would also violate all three inequalities (40), as was examined by [Sadjadpour and](#)

[Bhattacharya \(2007b\)](#). Finally, we have assumed isothermal conditions. Most experiments are not isothermal since the transformation from austenite to martensite is exothermic through the release of the latent heat. We can estimate the hardening in the other extremity of adiabatic conditions ( $q = r = 0$ ) following [Sadjadpour and Bhattacharya \(2007b\)](#). Under our constitutive assumption and the additional assumption that the dissipation due to kinetics is small compared to latent heat, (30) reduces to

$$c_p \dot{\theta} = \frac{L}{\theta_c} \dot{\lambda}. \quad (42)$$

We can integrate this from time  $t_1$  to  $t_2$  to obtain

$$\frac{\theta(t_2)}{\theta(t_1)} = \exp\left(\frac{L(\lambda(t_2) - \lambda(t_1))}{c_p \theta_c}\right), \quad (43)$$

or the temperature increases with advancing transformation. Returning now to the loading-unloading cycle under tension, the volume fraction is initially zero and the temperature is  $\theta_0$ . As we begin loading from the austenite, the temperature remains constant and things remain unchanged from before till we begin transformation. Thus, as before,

$$\sigma_{MS}^t = \frac{d_\lambda^c}{\varepsilon_i^t} + \frac{L(\theta_0 - \theta_c)}{\theta_c \varepsilon_i^t}. \quad (44)$$

As transformation proceeds, the temperature rises. However, the driving force has to remain at the critical value  $d_\lambda^c$ , and thus,

$$\sigma(t)\varepsilon_i^t - \frac{L}{\theta_c}(\theta(t) - \theta_c) = d_\lambda^c \Rightarrow \sigma(t) = \frac{d_\lambda^c}{\varepsilon_i^t} + \frac{L}{\varepsilon_i^t} \left( \frac{\theta_0}{\theta_c} \exp\left(\frac{L\lambda(t)}{c_p}\right) - 1 \right).$$

The transformation saturates at  $\lambda = \lambda^{t,c}$  given in (36); here,

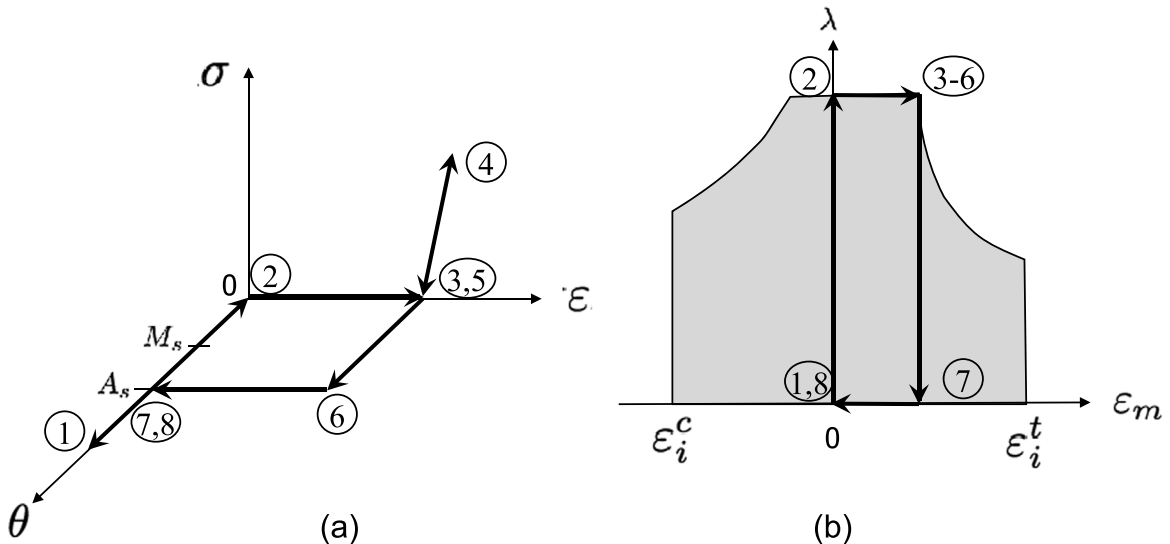
$$\sigma_{MF}^t = \frac{d_\lambda^c}{\varepsilon_i^t} + \frac{L}{\varepsilon_i^t} \left( \frac{\theta_0}{\theta_c} \exp\left(\frac{L\lambda^{t,c}}{\theta_c c_p}\right) - 1 \right). \quad (45)$$

Thus, hardening due to latent heat of transformation that would occur under adiabatic conditions is given by

$$\sigma_{MF}^{t,c} - \sigma_{MS}^{t,c} = \frac{L\theta_0}{\theta_c \varepsilon_i^{t,c}} \left( \exp\left(\frac{L\lambda^{t,c}}{\theta_c c_p}\right) - 1 \right). \quad (46)$$

### 3.1.2. Shape-memory effect

We now subject our one dimensional specimen to a controlled (infinitely slow) temperature-load cycle to explore the shape-memory effect. We begin as before with an unloaded specimen at a temperature above  $A_f$  with  $\lambda = 0$  and  $\varepsilon_m = 0$ . This



**Fig. 3.** Shape-memory effect in one dimension. (a) Stress-strain-temperature curve and (b) the volume fraction-nominal transformation strain space.

is marked as the points 1 in Fig. 3. We lower the temperature below  $M_s$  and  $M_f$  so that the material transforms to martensite and  $\lambda = 1$  at points 2. We now keep the temperature constant and subject the material to a strain controlled loading unloading cycle. The moment that strain, and consequently the stress becomes positive, the nominal transformation strain  $\varepsilon_m$  takes the value  $\varepsilon_i^t$  to optimize the energy and the material traverses from points 2 to points 3 in the figure. Now, the effective transformation strain is the saturation transformation strain  $\varepsilon_s^t$ . Note that in Fig. 3, this event happens at a value of stress equal to  $d_m^c = 0$ ; non-zero values would raise the stress of the martensite reorientation plateau. Further loading takes the material to points 4, and unloading brings it to 5 (which coincides with 3 in the stress–strain–temperature space). We now keep the stress zero and heat the specimen. There is no change in strain or transformation as the material traverses from points 5 to 6. At points 6, the temperature has reached  $A_s$  so that  $d_i = \omega = d_i^-$  and the reverse transformation begins. As the material traverses from points 6 to points 7, the volume fraction goes from  $\lambda = 1$  to  $\lambda = 0$ . Once it reaches zero, the  $\varepsilon_m$  resets to 0 and the material is in the state indicated by points 8 in the figure. Further heating brings the material back to the starting points 1.

Note that in the thermal cycle, there is no restriction on the amount of transformation, and the specimen is able to transform fully to martensite. Further, rate and heat transfer effects would make the curves rounded.

### 3.1.3. Actuation

We now examine the behavior of the model for actuation by studying a constant stress, temperature cycle. In this subsection, we assume  $d_m^c > 0$  and  $f_{\lambda\lambda} > 0$ . This changes the transformation temperatures to

$$M_s = \theta_c - \frac{d_i^c \theta_c}{L}, \quad (47)$$

$$M_f = \theta_c - \frac{(d_i^c + f_{\lambda\lambda}) \theta_c}{L}, \quad (48)$$

$$A_s = \theta_c + \frac{(d_i^c - f_{\lambda\lambda}) \theta_c}{L}, \quad (49)$$

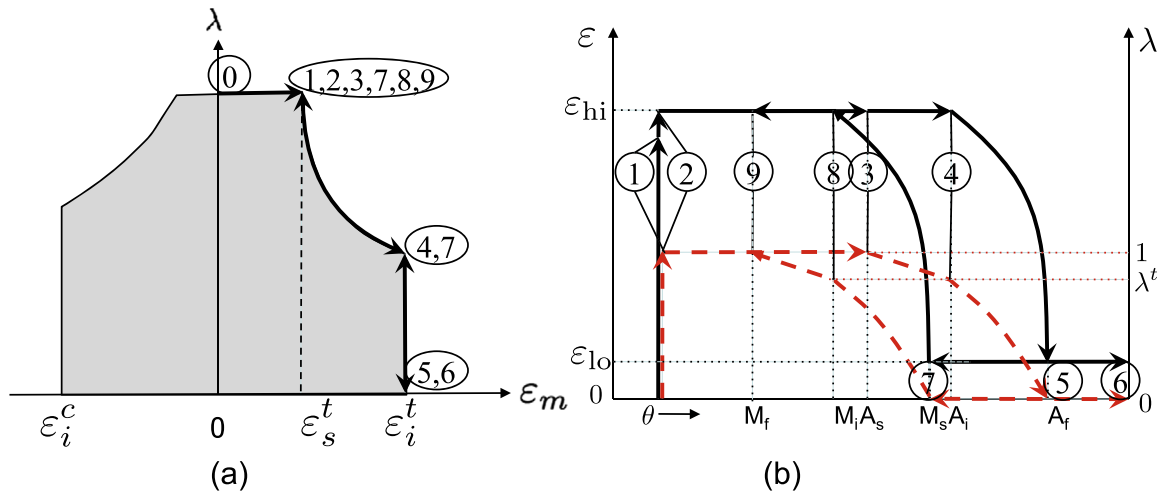
$$A_f = \theta_c + \frac{d_i^c \theta_c}{L}. \quad (50)$$

instead of (41).

Consider a material at a temperature  $\theta_0$  well below  $M_f$  with self-accommodated martensite at the points marked 0 in Fig. 4. We begin to apply a uniaxial stress to the material. As soon as it reaches the value

$$\sigma = \sigma_m = d_m^c, \quad (51)$$

the martensite reorients and the martensite strain  $\varepsilon_m$  jumps from zero to the value  $\varepsilon_s^t$ , i.e., to the points 1 in the figure. Since  $\lambda = 1$ , this is also the value of the effective transformation strain. We continue to increase the stress and  $\varepsilon_m$  remains at  $\varepsilon_s^t$  until the material reaches a final stress  $\sigma$  at points 2 in the figure. The value of the strain at this point is



**Fig. 4.** Actuation effect in one dimension. (a) The volume fraction-nominal transformation strain space. (b) Strain (black solid) and volume fraction (red dashed) as a function of temperature. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

$$\varepsilon = \varepsilon_{hi} = \varepsilon_s^t + \frac{\sigma}{E} \quad (52)$$

We now begin to heat the material. The driving force for transformation (19) is given by

$$d_\lambda = \sigma \varepsilon_s^t - L \frac{(\theta - \theta_c)}{\theta_c} - f_{\lambda\lambda}. \quad (53)$$

This force is initially positive but eventually becomes negative and reaches the critical value  $-d_\lambda^c$  when the temperature reaches

$$A_s(\sigma) = \theta_c + \frac{\theta_c}{L} (d_\lambda^c - f_{\lambda\lambda} + \sigma_m \varepsilon_s^t) \quad (54)$$

and the material reaches points 3 in Fig. 4. The martensite now begins to convert to austenite,  $\lambda$  begins to decrease as temperature increases and the material traverses from points 3 to 4 in the figure. Note that since the stress is still above  $\sigma_m$ , the nominal transformation strain  $\varepsilon_m$  increases to maintain the constraint  $\lambda \varepsilon_m = \varepsilon_s^t$ . However, notice that the strain remains at the value  $\varepsilon_{hi}$  in (52).<sup>4</sup> The driving force is

$$d_\lambda = \frac{\sigma \varepsilon_s^t}{\lambda} - L \frac{(\theta - \theta_c)}{\theta_c} - f_{\lambda\lambda} \lambda^{n_{\lambda\lambda}}. \quad (55)$$

In particular, the temperature continues to increase during reverse transformation to maintain a driving force value of  $-d_\lambda^c$ . The material reaches points 4 when the nominal transformation strain is  $\varepsilon_i^t$  and the volume fraction is  $\lambda_t$  given in (36) at a temperature

$$A_i(\sigma) = \theta_c + \frac{\theta_c}{L} \left( d_\lambda^c - f_{\lambda\lambda} \left( \frac{\varepsilon_s^t}{\varepsilon_i^t} \right)^{n_{\lambda\lambda}} + \sigma_m \varepsilon_i^t \right). \quad (56)$$

Further heating causes further reverse transformation as the material traverses from points 4 to 5 in Fig. 4. The nominal transformation strain remains at  $\varepsilon_i^t$ , but the effective transformation strain is  $\lambda \varepsilon_i^t$  so that the total strain  $\varepsilon = \lambda \varepsilon_i^t + \sigma/E$  begins to decrease with  $\lambda$ . The driving force is now

$$d_\lambda = \sigma \varepsilon_i^t - L \frac{(\theta - \theta_c)}{\theta_c} - f_{\lambda\lambda} \lambda^{n_{\lambda\lambda}}, \quad (57)$$

and the material needs further heating to sustain reverse transformation. It reaches points 5 when  $\lambda = 0$ , the temperature

$$A_f(\sigma) = \theta_c + \frac{\theta_c}{L} (d_\lambda^c + \sigma \varepsilon_i^t) \quad (58)$$

and strain

$$\varepsilon_{lo} = \frac{\sigma}{E} \quad (59)$$

Further heating does not change the state or the strain.

The material reaches the maximum temperature at points 6, and then begins to cool. The driving force  $d_\lambda = \sigma \varepsilon_i^t - L \frac{(\theta - \theta_c)}{\theta_c}$  is negative; however since  $\lambda = 0$ , further reverse transformation does not occur. As the temperature decreases, the driving force becomes positive and eventually reaches the value  $d_\lambda^c$  at temperature

$$M_s(\sigma) = \theta_c + \frac{\theta_c}{L} (-d_\lambda^c + \sigma_m \varepsilon_i^t) \quad (60)$$

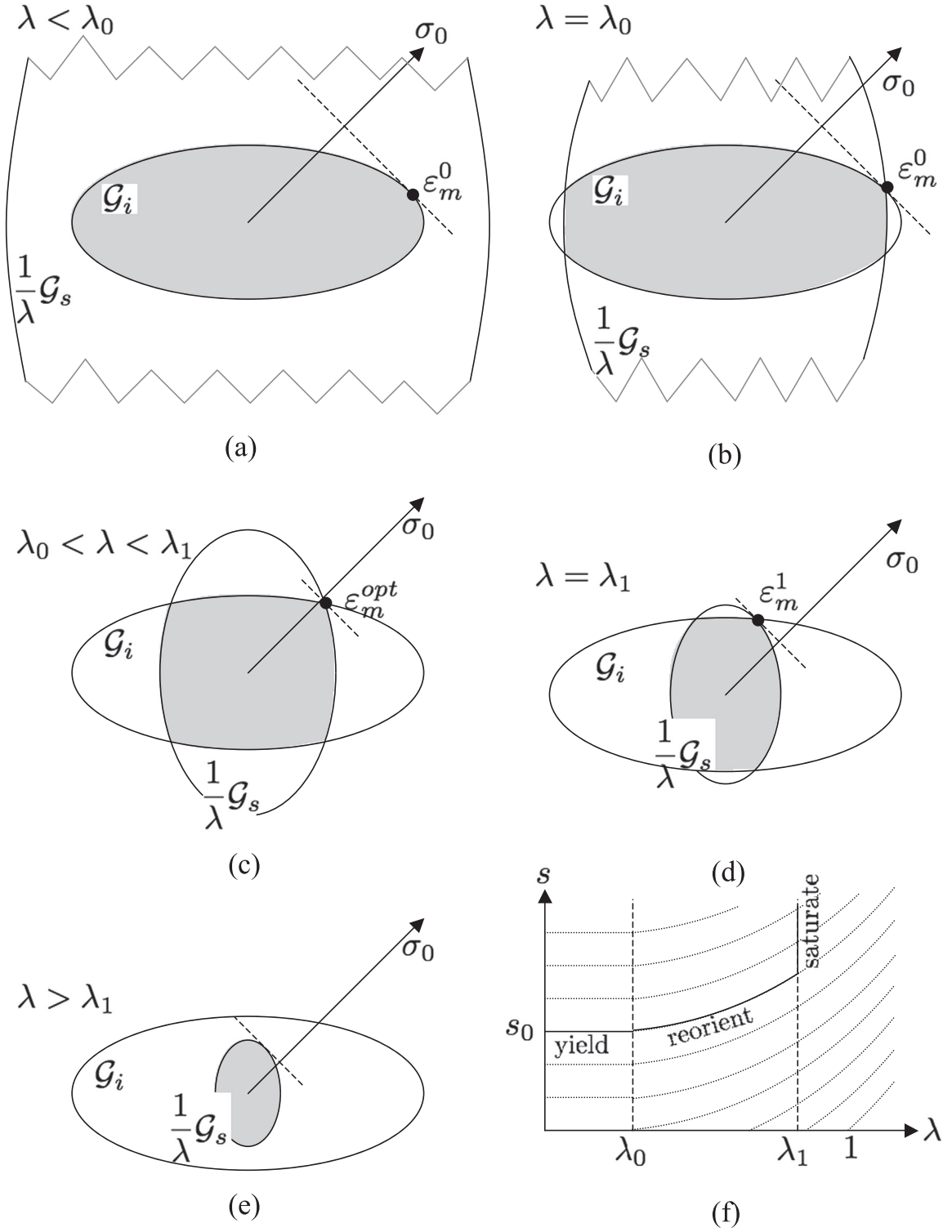
indicated as points 7. The transformation from austenite to martensite now begins and  $\lambda$  begins to increase along with the strain as the material traverses from points 7 to 8 with decreasing temperature. The driving force now is given by the expression (57), thus the material needs continued cooling to sustain transformation. It finally reaches the points 8 when the volume fraction reaches the value  $\lambda_t$  given in (36) and the temperature reaches the value

$$M_i(\sigma) = \theta_c + \frac{\theta_c}{L} \left( -d_\lambda^c - f_{\lambda\lambda} \left( \frac{\varepsilon_s^t}{\varepsilon_i^t} \right)^{n_{\lambda\lambda}} + \sigma_m \varepsilon_i^t \right). \quad (61)$$

The strain has now reached the value  $\varepsilon_{hi}$  in (52).

Further heating continues transformation;  $\lambda$  increases and the driving force is given by (55). The nominal transformation

<sup>4</sup> This decoupling between phase fraction and strain is a result of the simplifying assumptions in this analytic section including  $m_s = n_{\lambda\lambda} = \infty$ .



**Fig. 5.** The different eccentricities of the set  $\mathcal{G}_i$  and  $\mathcal{G}_s$  can lead to a significant reorientation of the martensite. (a)–(e) The optimal nominal transformation strain  $\epsilon_m$  for various volume fractions; the intersection  $\mathcal{G}_i \cap \frac{1}{\lambda}\mathcal{G}_s$  are shown in grey. (f) Level lines of the function  $f$  defined in (65).

strain  $\varepsilon_m$  decreases as  $\lambda$  increases to maintain the constraint  $\lambda \varepsilon_m = \varepsilon_s^t$  and the total strain remains fixed. The transformation is complete when  $\lambda = 1$  at a temperature equal to

$$M_f(\sigma) = \theta_c + \frac{\theta_c}{L} \left( -d_\lambda^c - f_{\lambda\lambda} + \sigma_m \varepsilon_s^t \right) \quad (62)$$

at points 9 in the figure. Continued cooling does not give any change in state or strain. The cycle repeats from points 2 in the figure if the material is reheated.

### 3.2. Multiple dimensions and reorientation of martensite

Having explored the one-dimensional version of the model, we return to the multidimensional setting. Recall the evolution equations that accounted for both driving forces and kinetic relations (28) and (29). According to these equations, but assuming the hard constraint (31), the rate-independent kinetics for  $\varepsilon_m$  dictate that at a given discretized time  $t$  (with the previous time step denoted  $t^-$ ),  $\varepsilon_m$  takes on values according to:

$$\varepsilon_m(t) = \begin{cases} \arg \max_{\xi \in \mathcal{G}_i \cap \frac{1}{\lambda} \mathcal{G}_s} \sigma(t): \xi \quad \sigma(t) \neq 0, \\ \varepsilon_m(t^-) & \sigma(t) = 0 \text{ \& } \lambda > 0, \\ 0 & \sigma(t) = 0 \text{ \& } \lambda = 0. \end{cases} \quad (63)$$

The kinetics for the  $\lambda$  may be described as follows. As long as  $\lambda \in (0, 1)$  and  $\lambda \varepsilon_m$  is in the interior of  $\mathcal{G}_s$ ,  $d_\lambda = \sigma: \varepsilon_m - \omega(\theta)$  and  $\lambda$  evolves as in the smooth case. Once we reach the boundary of either set,  $\lambda$  can only evolve in such a manner that keeps  $\lambda$  and  $\lambda \varepsilon_m$  inside the set (i.e., tangentially to the boundary or to the interior). For rate-independent kinetics, we obtain

$$-d_\lambda^c + \sigma: \varepsilon_m - \omega(\theta) - G_i \in \varepsilon_m: \partial G_s(\lambda \varepsilon_m) + \partial G(\lambda). \quad (64)$$

Now, consider an isothermal stress-controlled experiment where the temperature is held constant at a value significantly higher than the transformation temperature so that  $\omega(\theta) > d_\lambda^c > 0$  and the specimen is subject to a proportional load–unload cycle (i.e.,  $\sigma(t) = s(t)\sigma_0$  for a scalar  $s(t)$  that starts at zero, monotonically increases, peaks and then monotonically decreases to zero). The specimen is initially in the austenite state with  $\lambda = 0$  and  $\varepsilon_m = 0$ . We seek to simultaneously solve (28) and (29) for  $\lambda$  and  $\varepsilon_m$  at each  $t$ .

While the solution is similar to the one-dimensional situation described earlier, there are some differences since the interplay between the two constraints,  $\mathcal{G}_i$  and  $\mathcal{G}_s$ , is more complex in multiple dimensions. So we start by understanding this interplay by examining the solution of (63) for various  $\lambda$ . This is shown schematically in Fig. 5. When  $\lambda$  is extremely small,  $\frac{1}{\lambda} \mathcal{G}_s$  is extremely large and the optimization over  $\mathcal{G}_i \cap \frac{1}{\lambda} \mathcal{G}_s$  is equal to optimization over  $\mathcal{G}_i$  (see Fig. 5(a)). We call this optimal value  $\varepsilon_m^0$ , and note that it lies on the boundary of  $\mathcal{G}_i$  and is independent of  $\lambda$  for  $\lambda$  small enough. We define  $\lambda_0$  to be the largest value of  $\lambda$  for which this is true (see Fig. 5(b)):

$$\lambda_0 := \max \left\{ \lambda \in [0, 1]: \max_{\xi \in \mathcal{G}_i \cap \frac{1}{\lambda} \mathcal{G}_s} \sigma_0: \xi = \max_{\xi \in \mathcal{G}_i} \sigma_0: \xi \right\}.$$

As  $\lambda$  becomes larger than  $\lambda_0$ , both sets play a role in determining the optimal value  $\varepsilon_m$ , and this moves along the boundary of  $\mathcal{G}_i$  with increasing  $\lambda$  (see Fig. 5(c)). However, if  $\lambda$  becomes very large so that the set  $\frac{1}{\lambda} \mathcal{G}_s \subset \mathcal{G}_i$ , then the optimization over  $\mathcal{G}_i \cap \frac{1}{\lambda} \mathcal{G}_s$  is equal to optimization over  $\frac{1}{\lambda} \mathcal{G}_s$  (see Fig. 5(e)). We define  $\lambda_1$  to be the smallest value of  $\lambda$  for which this is true (see Fig. 5(d)):

$$\lambda_1 := \min \left\{ \lambda \in [0, 1]: \max_{\xi \in \mathcal{G}_i \cap \frac{1}{\lambda} \mathcal{G}_s} \sigma_0: \xi = \max_{\xi \in \frac{1}{\lambda} \mathcal{G}_s} \sigma_0: \xi \right\} = \min \left\{ \lambda \in [0, 1]: \max_{\xi \in \mathcal{G}_i \cap \frac{1}{\lambda} \mathcal{G}_s} \sigma_0: \xi = \max_{\xi \in \mathcal{G}_s} \sigma_0: \xi \right\}.$$

The second identity above is interesting. It says that the effective transformation strain  $\lambda \varepsilon_m$  has reached its maximum value consistent with its constraint at  $\lambda_1$ . In other words, the transformation saturates at this value.

With these definitions in place, we seek to simultaneously solve Eqs. (63) and (64) for  $\lambda$  and  $\varepsilon_m$  at each  $s$ . For  $\lambda \in (0, \lambda_1)$ , the constraint  $\lambda \varepsilon_m \in \mathcal{G}_s$  is not completely saturated and we may combine the two equations to obtain,

$$d_\lambda^c = \max_{\xi \in \mathcal{G}_i \cap \frac{1}{\lambda} \mathcal{G}_s} s \sigma_0: \xi - \omega =: f(s, \lambda). \quad (65)$$

Observe,

1. For any fixed  $\lambda$ ,  $f$  is monotone increasing in  $s$ .
2. For any fixed  $s$ ,  $f$  is monotone decreasing in  $\lambda$ .
3. For any fixed  $s$ ,  $f$  is constant for all  $\lambda \in [0, \lambda_0]$ .
4.  $f < 0$  for  $s$  small enough.

Together, (1)–(3) imply that in the  $\lambda$ – $s$  plane, the level curve of the function  $f$  is horizontal for  $\lambda \leq \lambda_0$  and monotone increasing beyond that as shown in Fig. 5(f). The level curve corresponding to  $f = d_\lambda^+$  defines the evolution of the  $\lambda$  with increasing  $s$ . The point 4 implies that this curve starts at a value  $s = s_0 > 0$ . The level curve  $f = d_\lambda^+$  determines the progress of the transformation. It starts at a finite value of stress, proceeds rapidly and saturates at  $\lambda_1$ . For each  $\lambda$ , we can obtain the corresponding nominal martensite strain  $\varepsilon_m$  from (63) and the overall transformation strain is given by  $\lambda \varepsilon_m$ . We now elaborate on this evolution.

### 3.2.1. Transformation yield

As  $s$  departs from zero,  $\varepsilon_m$  jumps according to (63) to the point  $\varepsilon_m^0$  in  $\mathcal{G}_i$  that has the maximum projection along the direction  $\sigma_0$  (see Fig. 5(a)). This is because  $\lambda = 0$  and  $\frac{1}{\lambda} \mathcal{G}_s$  is unbounded and  $\frac{1}{\lambda} \mathcal{G}_s \cap \mathcal{G}_i = \mathcal{G}_i$ . Since  $s$  is small,  $d_\lambda$  is negative, and  $\lambda$  is at its minimum value of 0. As  $s$  increases,  $\varepsilon_m$  remains unchanged, but  $d_\lambda$  increases monotonically till it becomes positive, and eventually reaches the value  $d_\lambda^c$  when

$$s(t) = s_0 = \frac{d_\lambda^c + \omega}{\sigma_0 \cdot \varepsilon_m^0} = \min_{\xi \in \mathcal{G}_i} \frac{d_\lambda^c + \omega}{\sigma_0 \cdot \xi}.$$

At this point, the transformation begins and proceeds rapidly as  $\lambda$  jumps from zero to a finite value  $\lambda_0$ . The stress of course remains constant at  $s_0$ , but the strain increases by  $\lambda_0 \varepsilon_m^0$ . Thus we may say that the material has yielded by transformation. Note that there is no hardening because we are in the isothermal, rate-independent setting, still assuming  $f_{\lambda\lambda} = 0$ .

Heuristically, as stress is applied to the material, favorably oriented variants of martensite nucleate in isolated grains that are favorably oriented with respect to the applied stress. The intergranular interactions remain limited as long as the volume fractions remain small. Thus, we have an apparent yield.

Following [Lexcelle et al. \(2002\)](#), we may define a transformation yield surface to be the locus of stress at which the material first yields by transformation. We have shown that the transformation yield surface is given by

$$Y = \left\{ \sigma : |\sigma^{dev}| = \min_{\xi \in \mathcal{G}_i} \frac{d_\lambda^c + \omega}{\hat{\sigma} \cdot \xi} \text{ where } |\sigma| = \sqrt{\sigma \cdot \sigma}, \hat{\sigma} = \frac{1}{|\sigma|} \sigma \right\}. \quad (66)$$

This is the boundary of the convex dual of the set of nominal transformation strains  $\mathcal{G}_i$ .

### 3.2.2. Hardening and reorientation of martensite

As  $s$  increases above  $s_0$ ,  $\lambda$  increases gradually beyond  $\lambda_0$  to solve (65). In particular, stress must increase to sustain continued transformation. Thus, we see an apparent hardening in the response despite the fact that we are in the rate-independent, isothermal setting. The reason becomes evident by examining the evolution of  $\varepsilon_m$ . Recall from the discussion above and from Fig. 5 that  $\varepsilon_m$  evolves from  $\varepsilon_m^0$  along the boundary of  $\mathcal{G}_i$  towards  $\varepsilon_m^1$  with increasing  $s$ . We may describe this as the reorientation of martensite.

Heuristically, following the apparent yield associated with the formation of well-oriented variants in the well-oriented grains, sustained transformation requires the participation of less-suitably oriented grains. Further, intergranular constraints become significant with increasing volume fraction. The result is that sustained transformation can proceed only with increasing stress (hardening), and the formation of locally suboptimal variants of martensite (reorientation).

Notice that the amount of hardening and the extent of reorientation depend on the difference between the eccentricity between the two sets  $\mathcal{G}_i$  and  $\mathcal{G}_s$  as well as the loading mode. Consider the situation with no difference in eccentricity wherein one is a scaled copy of the other:  $\mathcal{G}_i = \alpha \mathcal{G}_s$  for some  $\alpha > 1$ . In that case, it is easy to see from the definitions above that  $\lambda_0 = \lambda_1$  and  $\varepsilon_m^0 = \varepsilon_m^1$ . Thus, in the absence of any difference in eccentricity, the transformation saturates at yield with neither hardening nor martensite reorientation. Generally, we do not expect the two sets to be scaled versions of each other since initiation is driven by orientations of grains favorable to transform and their corresponding nominal martensite strains, while saturation is driven by polycrystalline constraints. Thus, we expect the former to be dominated by transformation surfaces similar to those expected by well-oriented single crystals, while we expect the latter to be dictated more by processing anisotropy, or texture of the polycrystal. Along these same lines or reasoning, [Bhattacharya and Schlömerkemper \(2004\)](#) have argued that the two sets would have significantly different eccentricities. Finally, hardening and reorientation have been widely observed.

An interesting consequence of the reorientation of martensite is that it is possible for some components of strain to decrease as stress increases. Notice that  $\varepsilon_m$  goes from  $\varepsilon_m^0$  to  $\varepsilon_m^1$ , and thus the effective transformation strain goes from  $\lambda_0 \varepsilon_m^0$  to  $\lambda_0 \varepsilon_m^1$ . Therefore it is entirely possible that some component decrease significantly enough to overcome the increase in elastic strain. This does not signal an instability since we have seen that the amount of transformation increases gradually and the component of strain along the applied stress  $\lambda \varepsilon_m \cdot \sigma_0$  increases monotonically. To see the latter, note that for any  $\lambda_0 \leq \lambda_a \leq \lambda_b \leq \lambda_1$ ,

$$\lambda_a \varepsilon_m^a \cdot \sigma_0 = \max_{\xi \in \lambda_a \mathcal{G}_i \cap \mathcal{G}_s} \xi \cdot \sigma_0 \leq \max_{\xi \in \lambda_b \mathcal{G}_i \cap \mathcal{G}_s} \xi \cdot \sigma_0 = \lambda_b \varepsilon_m^b \cdot \sigma_0.$$

Analogously, in strain-control experiments, the material may exhibit an apparent softening in some stress–strain curves, but



again these phenomena do not indicate instability.

### 3.2.3. Transformation saturation

The transformation saturates when  $\lambda = \lambda_1$  and

$$s(t) = s_1 = \frac{d_\lambda^c + \omega}{\sigma_0 : \epsilon_m^1} = \min_{\xi \in \frac{1}{\lambda_1} \mathcal{G}_s} \frac{d_\lambda^c + \omega}{\sigma_0 : \xi}.$$

As the stress is increased beyond  $s_1$ , there is no further transformation and the material responds elastically with fixed transformation strain and retained austenite.

### 3.2.4. Failure of resolved stress criterion

It is common to use a resolved-stress criterion that states that the product of the stress at the onset of transformation times the component of strain at the saturation of transformation is a constant independent of the direction (mode) of applied stress. In our model, we see that this product,

$$s_0(\lambda_1 \epsilon_m^1 : \sigma_0) = (d_\lambda + \omega) \frac{\max_{\xi \in \frac{1}{\lambda_1} \mathcal{G}_s} \sigma_0 : \xi}{\max_{\xi \in \mathcal{G}_i} \sigma_0 : \xi}$$

generally depends on the load orientation. Thus the resolved stress criterion fails in our model, in agreement with experimental observations (Daly et al., 2008).

### 3.2.5. Unloading

As the applied stress peaks and begins to reduce, the material responds elastically till the stress reduces to

$$s(t) = s_2 = \frac{-d_\lambda^c + \omega}{\sigma_0 : \epsilon_m^1}.$$

Note that this quantity is strictly smaller than  $s_1$  because of the sign of  $d_\lambda^c$ . The reverse transformation begins at this stress and proceeds gradually with decreasing stress till the volume fraction reaches the value  $\lambda_0$  at

$$s(t) = s_3 = \frac{-d_\lambda^c + \omega}{\sigma_0 : \epsilon_m^0}.$$

At this stress, the reverse transformation proceeds rapidly to complete austenite at constant stress. Subsequently, the material unloads elastically to the origin.

## 3.3. Uniaxial tension and compression of an isotropic material

In this and the following subsection, we demonstrate how the reorientation and hardening can depend on loading mode. We focus on an isotropic material ( $C_i = C_s = \mathcal{I}$ ) and continue to use the hard constraint (31) for saturation. We show in this subsection that applied uniaxial stress leads to a response with no reorientation, very similar to the one-dimensional situation discussed in Section 3.1. In contrast, we will show in the following subsection that applied pure shear can lead to reorientation as described in Section 3.2.

Consider a specimen subjected to an uniaxial tensile or compressive stress  $\sigma = \text{diag}[0, 0, s]$  where  $\text{diag}[a, b, c]$  denotes a diagonal matrix with diagonal entries  $a, b$  and  $c$ . As discussed in the previous section, specifically (63), the nominal strain  $\epsilon_m$  is the value that maximizes the projection along the applied stress amongst those values that satisfy the constraint  $\epsilon_m \in \mathcal{G}_i \cap \frac{1}{\lambda} \mathcal{G}_s$ . Further, the nominal transformation strain  $\epsilon_m^i$  (respectively  $\epsilon_m^s$ ) that initiates (respectively saturates) the transformation is the value that maximizes the projection along the applied stress of points in  $\mathcal{G}_i$  (respectively  $\mathcal{G}_s$ ). Now, since the sets  $\mathcal{G}_\alpha = \{\xi : \text{tr } \xi = 0, g_\alpha(\xi) \leq 0\}$ , and the functions  $g_\alpha$  are differentiable away from the origin, the stress is normal to the level set of  $g_\alpha$  at  $\epsilon_m^\alpha$ :

$$\sigma^{\text{dev}} = c_\alpha \left( \frac{\partial g_\alpha}{\partial \xi} \bigg|_{\epsilon_m^\alpha} \right)^{\text{dev}} \quad (67)$$

for some scalar  $c_\alpha$ . Further from (7) and the fact that  $C_\alpha = \mathcal{I}$ ,

$$g_\alpha(\xi) = \sum_{n=1}^3 f_\alpha(\mu_n(\xi)) \Rightarrow \frac{\partial g_\alpha}{\partial \xi} = \sum_{n=1}^3 f'_\alpha(\mu_n(\xi)) e_n(\xi) \otimes e_n(\xi) \quad (68)$$

where  $f_\alpha(\zeta) = (|\zeta|^2 - a_{1\alpha})^{a_{2\alpha}} - b_\alpha^{a_{2\alpha}}/3$  and  $e_n(\xi)$  denotes the eigenvector of  $\xi$  corresponding to eigenvalue  $\mu_n(\xi)$  (see the Appendix for the details of differentiating eigenvalues and eigenvectors). For our choice of constants  $a_{1\alpha}$  and  $a_{2\alpha}$ ,  $f'_\alpha$  is strictly

monotone and extends from  $-\infty$  to  $\infty$  so that  $x = f'_\alpha(y)$  has a unique solution  $y = y(x)$ .

Thus, it follows from (67) that the eigenvectors of  $\varepsilon_m^\alpha$  are parallel to the principal axes of loading and that

$$\varepsilon_m^\alpha = (\text{diag}[b^\alpha, a^\alpha])^{\text{dev}} = \frac{2}{3}(a^\alpha - b^\alpha) \text{diag}\left[-\frac{1}{2}, -\frac{1}{2}, 1\right]. \quad (69)$$

Therefore, the nominal transformation strain for initiation and saturation,  $\varepsilon_m^i$  and  $\varepsilon_m^s$ , are parallel to each other and there is no reorientation of martensite.

Further, recalling the assumption that the material is elastically isotropic, the total strain and the nominal transformation strain have the following form:

$$\varepsilon = xI + y\varepsilon^0, \quad \varepsilon_m = z\varepsilon^0 \quad (70)$$

where  $I$  is the identity and  $\varepsilon^0 = \text{diag}\left[-\frac{1}{2}, -\frac{1}{2}, 1\right]$ . Because  $\sigma_{11} = \sigma_{22} = 0$ ,

$$x = \frac{\mu}{3\kappa}(y - \lambda z); \quad (71)$$

so

$$\sigma = \mu(y - \lambda z)(I + \varepsilon^0) = \frac{E}{2(1 + \nu)}(y - \lambda z)(I + \varepsilon^0). \quad (72)$$

Since we expect  $\sigma_{33} = E(\varepsilon_{33} - \lambda(\varepsilon_m)_{33})$  in uniaxial loading, we conclude that the uniaxial strain and nominal transformation strain are  $(1 + \nu)y$  and  $(1 + \nu)z$  respectively.

Now, substituting (70) into (7) gives

$$g_\alpha(\varepsilon_m) = 2 \left( \left| -\frac{1}{2}z \right| + a_{1\alpha} \frac{1}{2}z \right)^{a_{2\alpha}} + (|z| - a_{1\alpha}z)^{a_{2\alpha}} - b_\alpha^{a_{2\alpha}} = \begin{cases} (z^{1-a_{2\alpha}}(1 + a_{1\alpha})^{a_{2\alpha}} + (1 - a_{1\alpha})^{a_{2\alpha}})z^{a_{2\alpha}} - b_\alpha^{a_{2\alpha}} & z > 0, \\ (z^{1-a_{2\alpha}}(-1 + a_{1\alpha})^{a_{2\alpha}} + (-1 - a_{1\alpha})^{a_{2\alpha}})z^{a_{2\alpha}} - b_\alpha^{a_{2\alpha}} & z < 0. \end{cases} \quad (73)$$

Setting  $g_\alpha = 0$ , solving for  $z$  and recalling that the nominal transformation strain is  $(1 + \nu)z$ , we obtain

$$\varepsilon_i^c = \frac{(1 + \nu)b_i^{a_{2i}}}{2^{1-a_{2i}}(-1 + a_{1i})^{a_{2i}} + (-1 - a_{1i})^{a_{2i}}}, \quad (74)$$

$$\varepsilon_i^t = \frac{(1 + \nu)b_i^{a_{2i}}}{2^{1-a_{2i}}(1 + a_{1i})^{a_{2i}} + (1 - a_{1i})^{a_{2i}}}, \quad (75)$$

$$\varepsilon_s^c = \frac{(1 + \nu)b_s^{a_{2s}}}{2^{1-a_{2s}}(-1 + a_{1s})^{a_{2s}} + (-1 - a_{1s})^{a_{2s}}}, \quad (76)$$

$$\varepsilon_s^t = \frac{(1 + \nu)b_s^{a_{2s}}}{2^{1-a_{2s}}(1 + a_{1s})^{a_{2s}} + (1 - a_{1s})^{a_{2s}}}. \quad (77)$$

In summary, uniaxial tension and compression in an isotropic material is similar to the one-dimensional model presented in Section 3.1, and all the formulas there hold with  $\varepsilon_{i,s}^{t,c}$  as defined above. In particular, there is no reorientation of martensite. These mechanics are independent of the tension–compression asymmetry.

### 3.4. Shear (torsion) of an isotropic material

The situation is different in the case of shear or torsion of an isotropic material when one has tension–compression asymmetry.

Consider a material subjected to a pure shear stress

$$\sigma = s \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}. \quad (78)$$

It is easier to change frame to the principal frame of loading  $e_1, \frac{1}{\sqrt{2}}(e_2 - e_3), \frac{1}{\sqrt{2}}(e_2 + e_3)$  and where the stress is

$$\sigma = s \text{diag}[0, -1, 1]. \quad (79)$$

We argue as before, and obtain the nominal transformation strain at initiation and saturation by solving (67). It follows that the eigenvectors of  $\varepsilon_m^\alpha$  are parallel to the principal axes of loading and that

$$\varepsilon_m^\alpha = (\text{diag}[0, a^\alpha, b^\alpha])^{\text{dev}} = \frac{1}{3} \text{diag}[a^\alpha - b^\alpha, 2a^\alpha - b^\alpha, 2b^\alpha - a^\alpha]. \quad (80)$$

Observe that if  $f^\alpha$  is an even function, or if equivalently  $a_{1\alpha} = 0$ , or equivalently the material displays tension–compression

symmetry, then  $a^\alpha = -b^\alpha$  and  $\varepsilon^m = z \text{ diag}[0, 1, -1]$ . Thus, when the material has tension–compression symmetry, the response to a pure shear applied stress is a pure shear strain and there is no reorientation of martensite. So shear of a symmetric material behaves like the one-dimensional model in Section 3.1 and we could proceed as in the previous section and obtain formulas for  $\varepsilon_{i,s}^s$ .

However, when we have tension–compression asymmetry, which is natural in shape-memory alloys,  $f^\alpha$  is not an even function and  $a^\alpha \neq -b^\alpha$ . This has immediate consequences. First, it means that we see the *Poynting effect* where we have elongation/contraction in the direction perpendicular to the plane of shear (Poynting, 1909). Second, in general  $\frac{d^j}{b^j} \neq \frac{a^s}{b^s}$  and therefore the value of the nominal transformation strain at initiation and saturation will not be proportional to each other. Thus, we will have reorientation of martensite and hardening as discussed in Section 3.2. The algebra required for explicit analysis is very tedious beyond this point. Thus we proceed to provide more complicated situations with numerical examples after discussing the numerical implementation.

#### 4. Time discretization and numerical implementation

We seek to solve the equilibrium equation (13), heat equation (30), and the evolution equations (28) and (29) simultaneously. One approach is to use a staggered thermoelastic predictor – inelastic corrector approach. Given  $\varepsilon^n$ ,  $\theta^n$ ,  $\varepsilon_m^n$  and  $\lambda^n$  at the  $n$ th time step, we solve (13) and the heat equation (30) to obtain updated  $\varepsilon^{n+1}$  and  $\theta^{n+1}$ . In doing so, we regard the last three terms of the heat equation as a source. This is the thermoelastic predictor.

It remains to specify the inelastic corrector using the evolution equations (28) and (29). This is tricky because it involves constraints and non-smooth functions. Thus, we evoke Lagrange multipliers ( $\Lambda$ ). We introduce  $\Lambda_{tm}$  to satisfy the constraint  $\text{tr}(\varepsilon_m) = 0$ ,  $\Lambda_i$  to satisfy the constraint  $\varepsilon_m \in G_i$  or equivalently  $g_i \leq 0$ . We also introduce  $\Lambda_\lambda$  to satisfy the constraints  $0 \leq \lambda \leq 1$ . To distinguish between the direction and the magnitude of the nominal transformation strain increment, we discretize the martensite strain according to

$$\varepsilon_m^{n+1} = \varepsilon_m^n + \alpha_m D_m, \quad (81)$$

where  $D_m$  satisfies

$$D_m: D_m = 1. \quad (82)$$

We now rewrite the rate-independent evolution equations (28) and (29) as

$$\begin{aligned} \frac{1}{2}(\varepsilon^{n+1} - \lambda^{n+1} \varepsilon_m^{n+1}): \frac{\partial C}{\partial \lambda} \Big|_{\lambda^{n+1}} : (\varepsilon^{n+1} - \lambda^{n+1} \varepsilon_m^{n+1}) - \frac{dc_p}{d\lambda} \Big|_{\lambda^{n+1}} \theta^{n+1} \ln \left( \frac{\theta^{n+1}}{\theta_c} \right) - \sigma^{n+1}: \varepsilon_m^{n+1} \\ + L \frac{(\theta^{n+1} - \theta_c)}{\theta_c} + \varepsilon_m^{n+1}: \frac{\partial G_s}{\partial \xi} \Big|_{\lambda^{n+1} \varepsilon_m^{n+1}} + I(\lambda^{n+1}) \Lambda_\lambda + \frac{\partial g_\lambda}{\partial \lambda} \Big|_{\lambda^{n+1}} + \text{sign}(\lambda^{n+1} - \lambda^n) d_\lambda^c = 0, \end{aligned} \quad (83)$$

$$-\sigma^{n+1} + \Lambda_i \frac{\partial g_i}{\partial \xi} \Big|_{\varepsilon_m^{n+1}} + \frac{\partial G_s}{\partial \xi} \Big|_{\lambda^{n+1} \varepsilon_m^{n+1}} + \Lambda_{tm} I + d_m^c D_m = 0, \quad (84)$$

$$g_i(\varepsilon_m^{n+1}) = 0, \quad (85)$$

$$0 \leq \lambda^{n+1} \leq 1, \quad (86)$$

$$\text{tr}(\varepsilon_m^{n+1}) = 0, \quad (87)$$

$$D_m: D_m = 1, \quad (88)$$

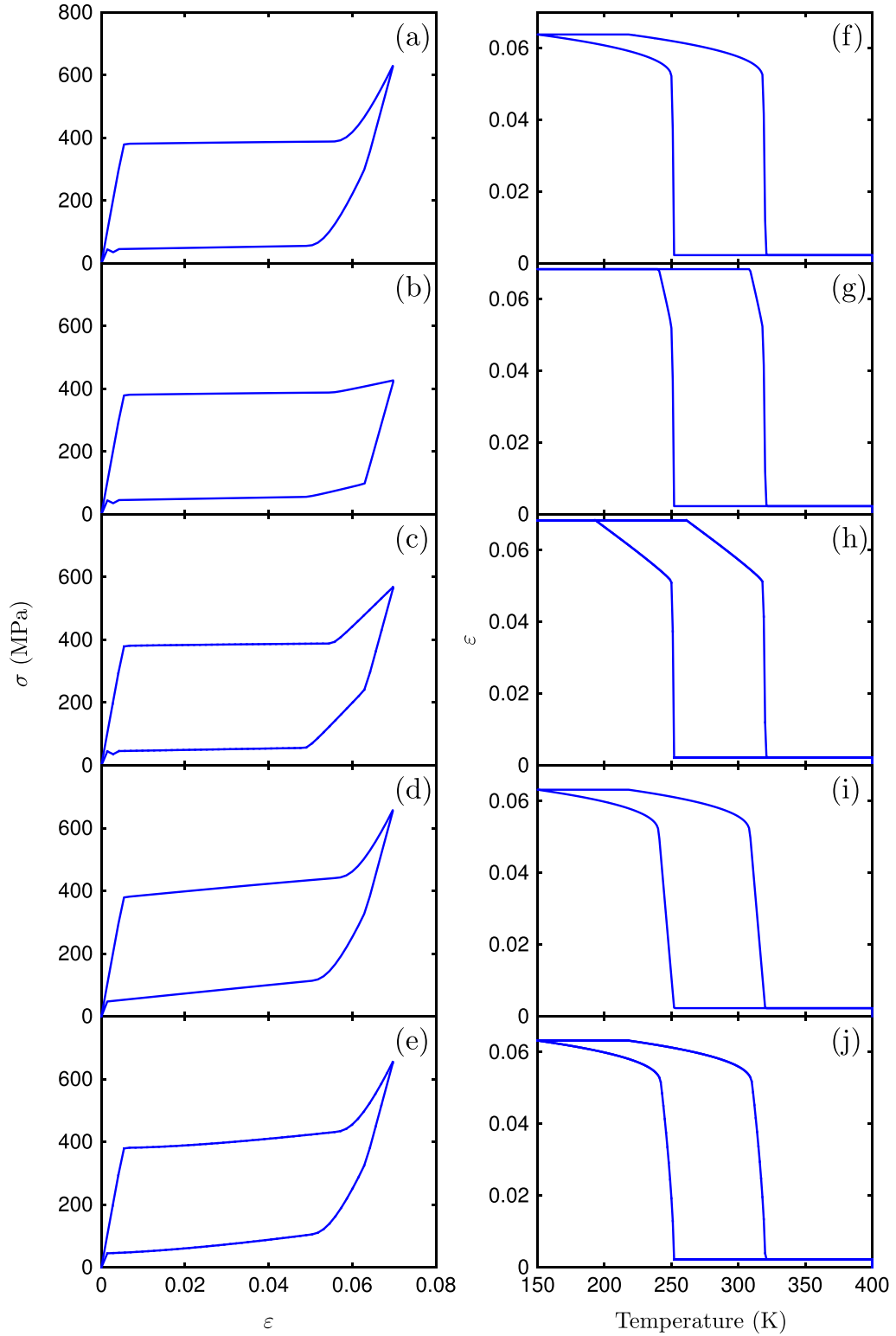
where

$$I(\lambda^{n+1}) = \begin{cases} 0 & 0 < \lambda < 1 \\ 1 & \text{else} \end{cases} \quad (89)$$

$$\sigma^{n+1} = C(\lambda^{n+1}): (\varepsilon^{n+1} - \lambda^{n+1} \varepsilon_m^{n+1}). \quad (90)$$

This gives 6 + 1 + 4 equations for 6 + 1 + 4 unknowns  $\{D_m, \lambda^{n+1}, \Lambda_i, \Lambda_{tm}, \alpha_m\}$ .

We note that Eqs. (83) (through 88) are carried out at each material point. Therefore these equations may be specified as a user defined material within a commercial finite element solver.



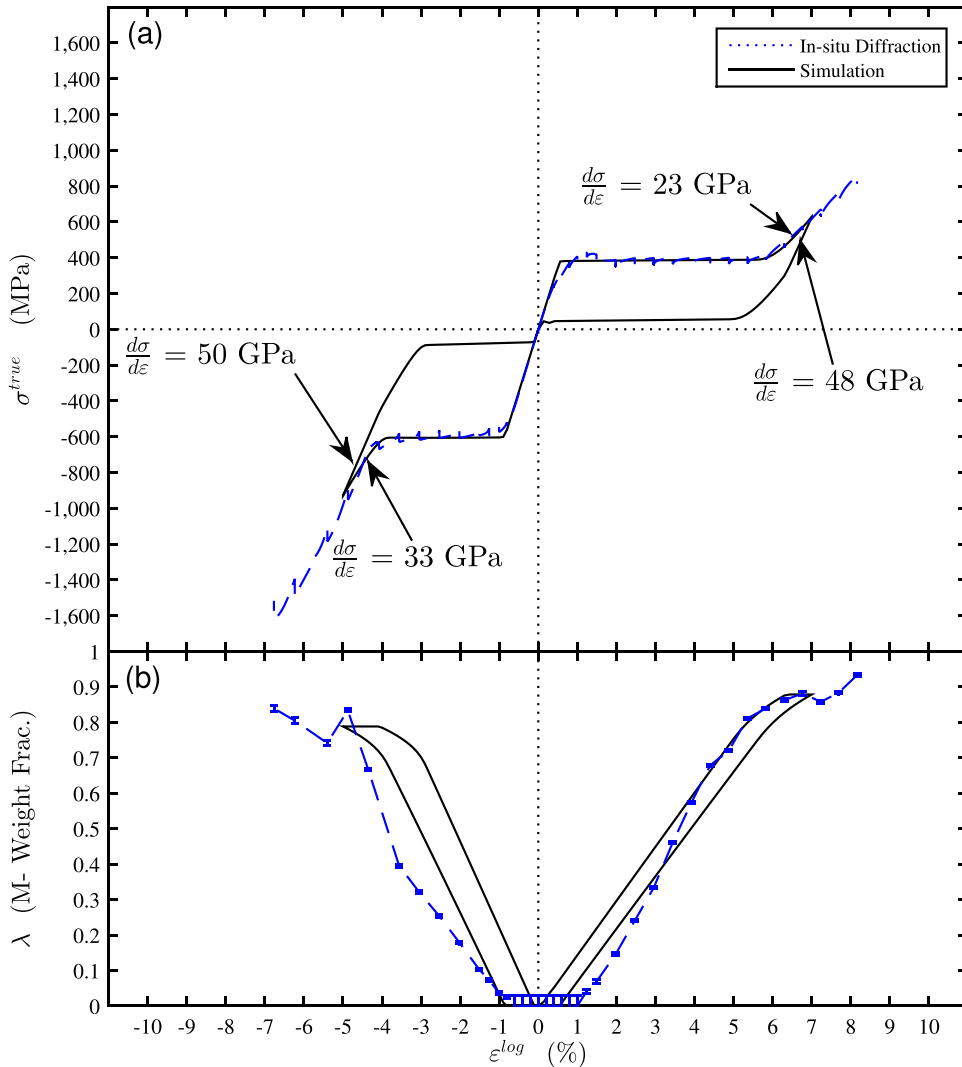
**Fig. 6.** (a)–(e) Superelastic and (f)–(j) actuation responses for various parameters. (a,f) Results for the parameters in Table 1. (b,g)  $m_s=3$ , rest as in Table 1. (c,h)  $m_s=3$ ,  $k_s=0.05$ , rest as in Table 1. Increasing  $k_s$  increases the slope of the post-apparent saturation hardening. (d,i)  $f_{\lambda\lambda}=6$ , rest as in Table 1. (e,j)  $f_{\lambda\lambda}=6$ ,  $n_{\lambda\lambda}=1.5$ , rest as in Table 1. The changes in simulated behaviors are discussed in detail in the text.

**Table 1**Parameters used as the baseline for the parameter study and also to fit the in-situ neutron diffractions experiments of [Stebner et al. \(2015\)](#).

Material Properties			Model Parameters		
$A_f = 293$ K	$A_s = 274$ K	$M_s = 225$ K	$a_{1i} = 0.25$	$a_{2i} = 10$	$b_i = 0.05$
$M_f = 195$ K	$c_p^A = 2.79 \frac{\text{MJ}}{\text{m}^3}$	$c_p^M = 2.79 \frac{\text{MJ}}{\text{m}^3}$	$a_{1s} = 0.33$	$a_{2s} = 10$	$b_s = 0.033$
$L = 100 \frac{\text{MJ}}{\text{m}^3}$	$d_\lambda^c = 11.0 \frac{\text{MJ}}{\text{m}^3}$	$d_m^c = 50$ MPa	$\theta_c = \frac{A_f + M_s}{2}$	$k_s = 0.008$	$m_s = 5$
$E^A = 70$ GPa	$E^M = 45$ GPa	$\nu^A = 0.45$	$f_{\lambda\lambda} = 1$	$n_{\lambda\lambda} = 1$	
		$\nu^M = 0.468$	$C_i = C_s = I$		

## 5. Model demonstration

We now demonstrate the abilities and limitations of the current model. We first conduct a parameter study followed by comparison of the model with two sets of experiments. In calibrating the model to experiments, we use a manual trial and error approach. Specifically, these calibrations are not optimized; instead the main purpose is to demonstrate the ability of the model to describe the complex mechanics of SMAs. Further, we consider simple tests of the material like uniform



**Fig. 7.** Comparison with in-situ neutron diffraction tests of [Stebner et al. \(2015\)](#). The model captures the tension–compression asymmetry as well as the fact that the transformation is does not complete, even on apparent saturation. (For interpretation of the references to color in the text, the reader is referred to the web version of this paper.)

**Table 2**

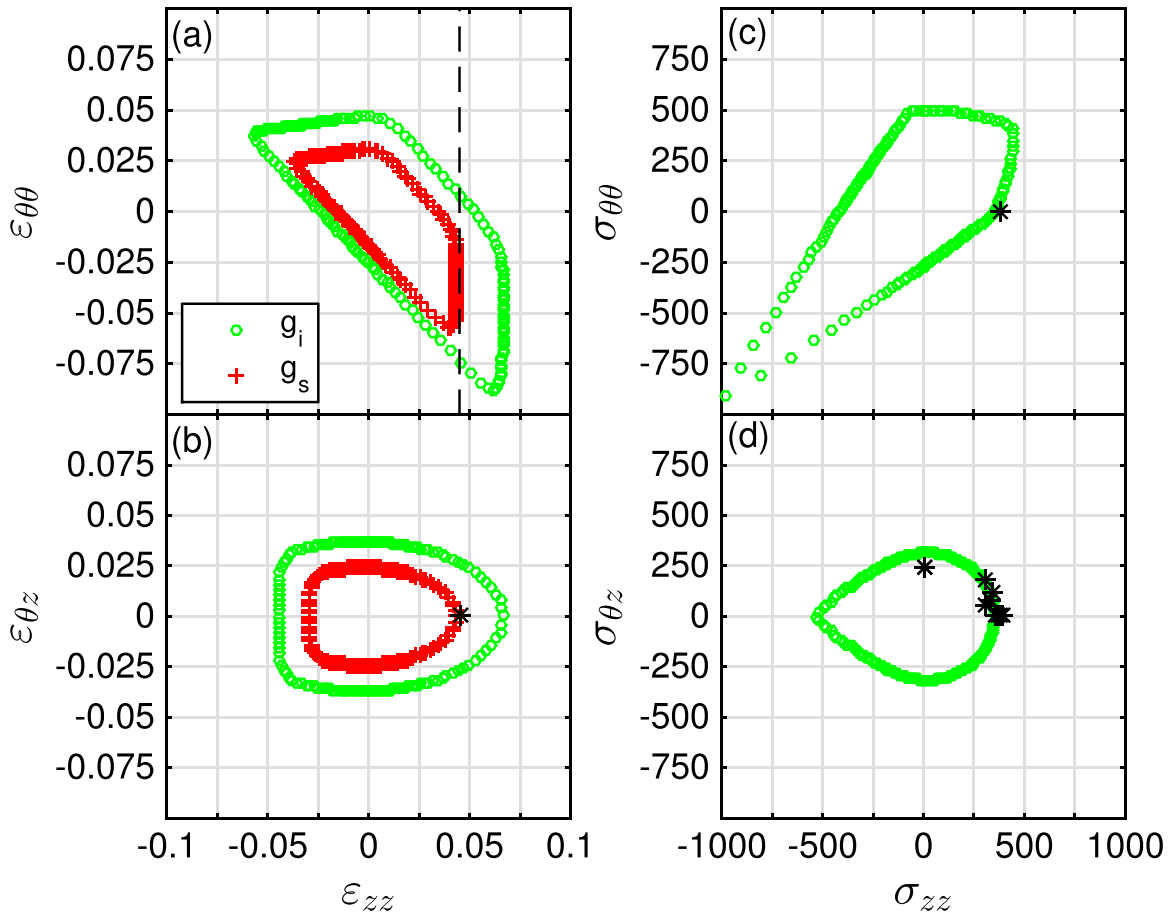
Parameters use to fit the multiaxial loading experiments of McNaney et al. (2003).

Material Properties			Model Parameters		
$A_f = 291.1$ K	$A_s = 266.6$ K	$M_s = 221.4$ K	$a_{1i} = -0.38$	$a_{2i} = 10$	$b_i = 0.046$
$M_f = 185.6$ K	$c_p^A = 2.79 \frac{\text{MJ}}{\text{m}^3}$	$c_p^M = 2.79 \frac{\text{MJ}}{\text{m}^3}$	$a_{1s} = -0.38$	$a_{2s} = 10$	$b_s = 0.03$
$L = 107 \frac{\text{MJ}}{\text{m}^3}$	$d_\lambda^c = 7.28 \frac{\text{MJ}}{\text{m}^3}$	$d_m^c = 50$ MPa	$\theta_c = \frac{A_f + M_s}{2}$	$k_s = 0.01$	$m_s = 5$
$E^A = 50$ GPa	$E^M = 22.6$ GPa	$\nu^A = -0.06$	$f_{\lambda\lambda} = 1$	$n_{\lambda\lambda} = 1$	$C_i = C_s = C$
		$\nu^M = 0.24$	$c_{11} = 1.6$	$c_{12} = 0.3$	$c_{13} = 0.2$
			$c_{22} = 1$	$c_{23} = 0.2$	$c_{33} = 0.7$
			$c_{44} = 0.9$	$c_{55} = 1$	$c_{66} = 1$

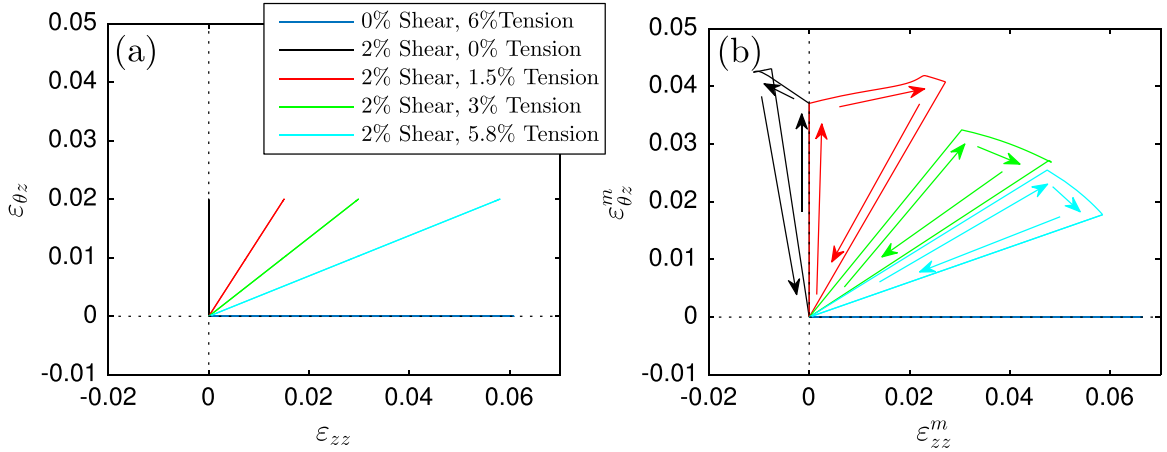
uniaxial and shear loading with specified temperature. Therefore, we do not have to solve the equilibrium equation (13) or the heat equation (30). Instead we only focus on the evolution laws (28) and (29) using the numerical discretization described in Eqs. (83) (through 88).

### 5.1. Parameter study

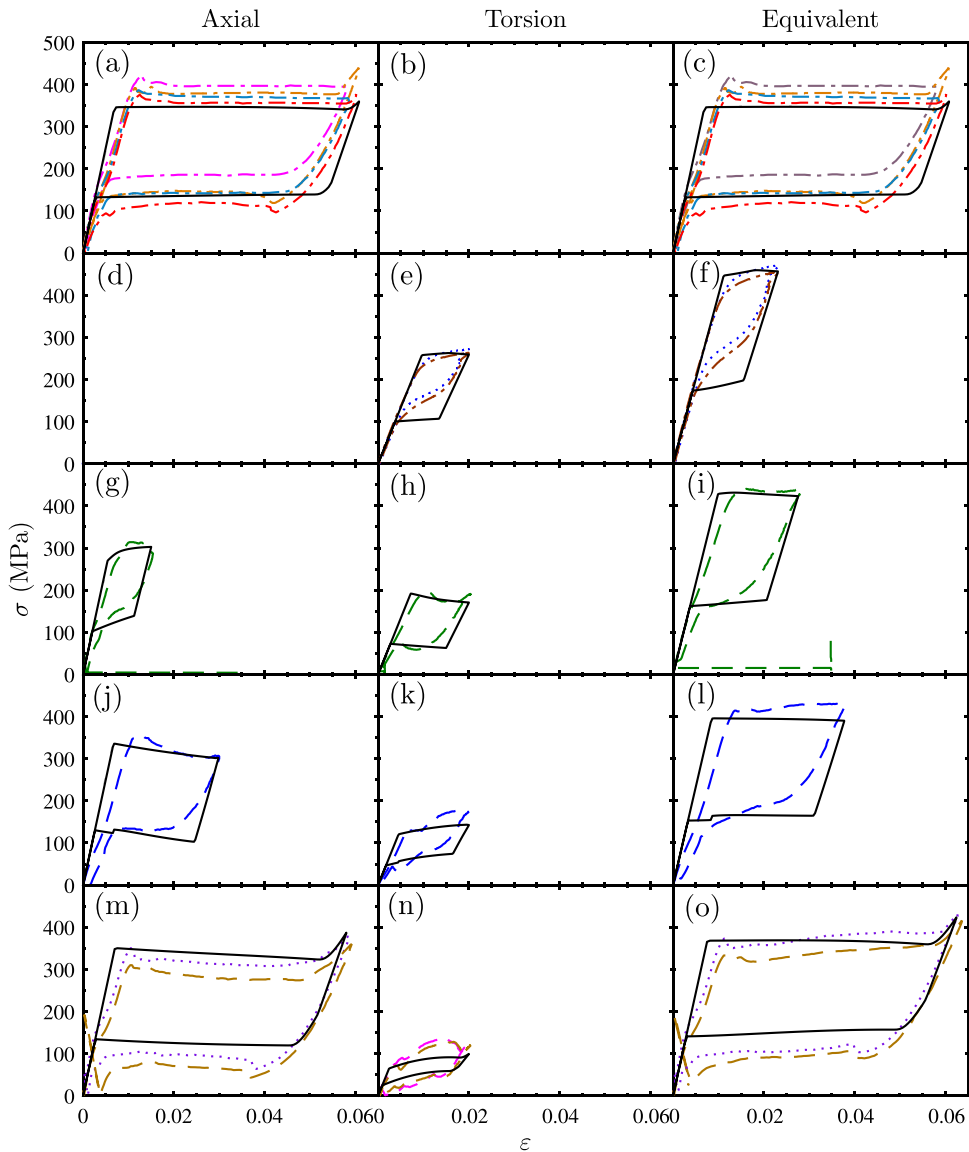
We begin with a brief study of model parameters that have not been documented before. The current model builds on the prior work of Sadjadpour and Bhattacharya (2007b), who used a free energy with the mechanical, latent heat, specific heat, and initiation terms of (3). They provided an extensive parameter study on initiation and kinetic parameters. We have replaced their invariant-based surface function (Sadjadpour and Bhattacharya, 2007b) with a principal value-based function following Cazacu et al. (2006), who provided a detailed study of the surface parameters. We do not repeat these parameter



**Fig. 8.** Level sets of the multi-axial calibration used to simulate the experiments of McNaney et al. are shown for (a)–(b) initiation and saturation strain surfaces and (c)–(d) the initiation stress surface. Data points extracted from the proportional loading experiments of McNaney (see Fig. 9) are indicated in black.

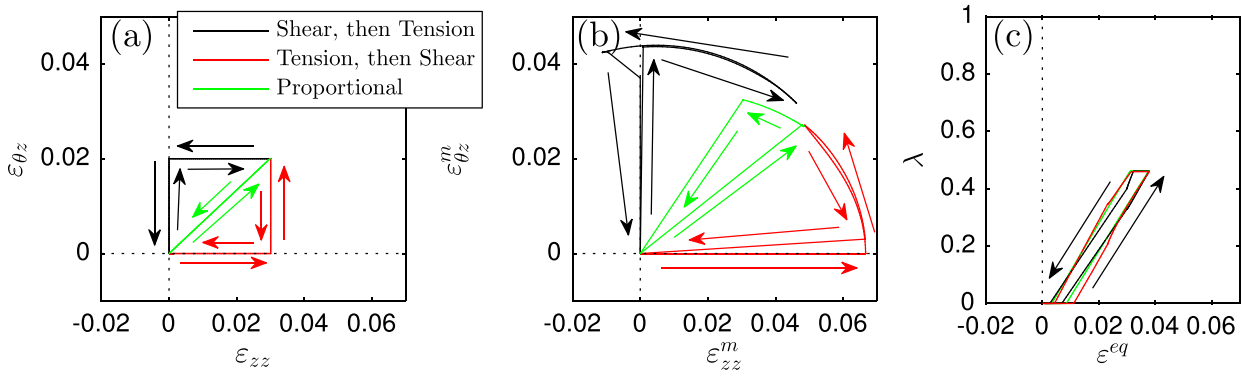


**Fig. 9.** (a) Total shear vs. axial strain imposed by McNaney et al. (2003) in proportional tests. (b) Corresponding nominal transformation strain paths. The stress–strain responses are shown in Fig. 10. (For interpretation of the references to color in the text, the reader is referred to the web version of this paper.)



**Fig. 10.** Comparisons of model simulations (black solid lines) to the proportional loading experimental data of McNaney et al. (broken colored lines; multiple sets indicate multiple experiments, and illustrate the experimental variation). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)





**Fig. 11.** (a) Three loading paths to the same total strain of 3% axial strain, 2% shear strain. (b) Corresponding evolution of the nominal transformation strain and (c) corresponding evolution of the volume fraction. The mechanical responses given in Figs. 10, 12 and 14.

studies. Finally, the interplay between the initiation and saturation was analytically explored in Section 3.2 for the case of rigid constraints. Therefore, we limit the ensuing numerical parameter study to the parameters  $f_{\lambda\lambda}$  and  $n_{\lambda\lambda}$  of the volume-fraction energy  $G_\lambda$  given in (6) as well as the parameters  $k_s$  and  $m_s$  of saturation energy  $G_s$  given in (10).

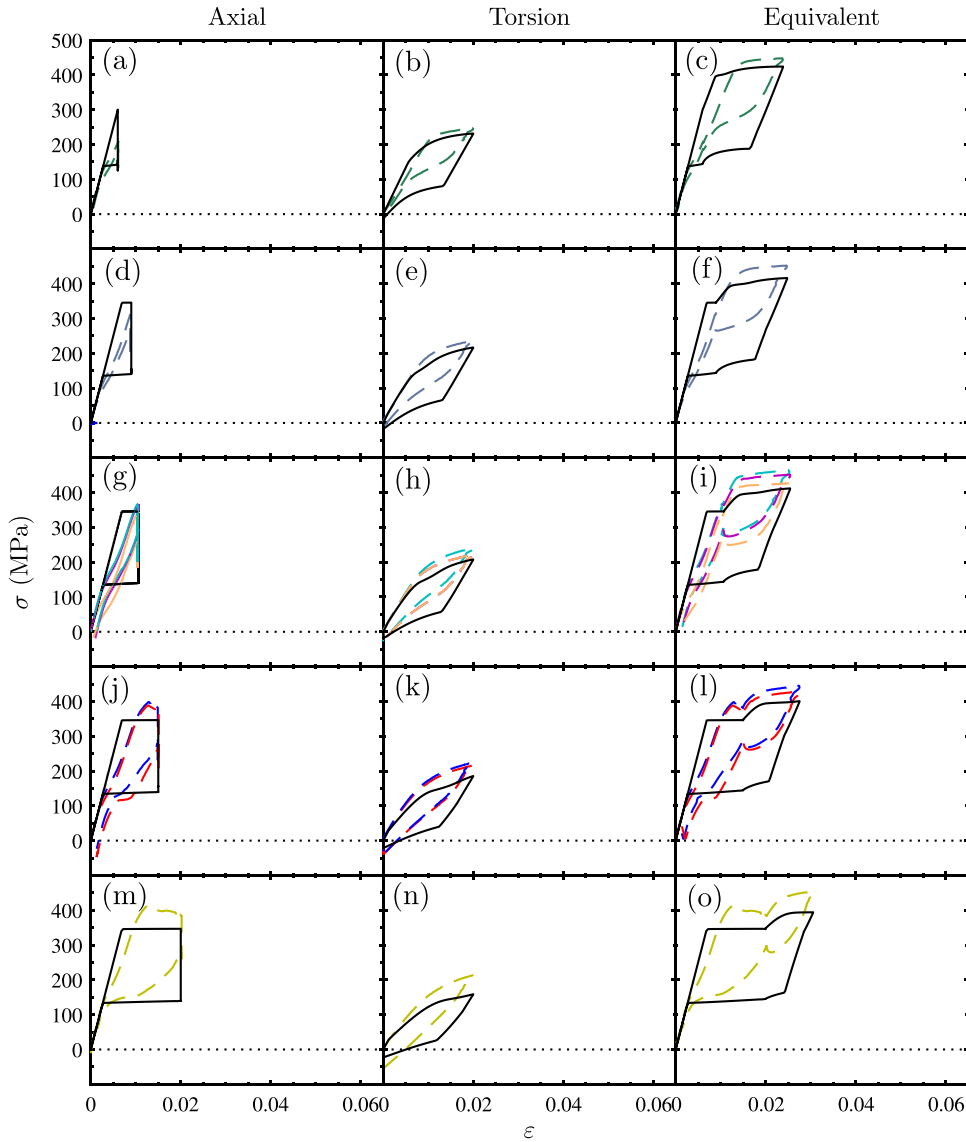
Fig. 6 shows the results of simulations of the superelastic responses of a material loaded to 7% strain and unloaded (a–e), as well as actuation responses of a material loaded to 160 MPa at 400 K, cooled to 150 K, and then heated back to 400 K (f–j) for various choices of these new parameters. Figs. 6(a,f) show the results for the base set of parameters in Table 1. As in the superelastic response of the previous section (Fig. 2) where we had an abrupt saturation criterion, the response is initially elastic followed by a plateau. However, after a plateau as in Fig. 2, the saturation is not abrupt, but begins smoothly. Further, the slopes of the loading and unloading portions at the highest stresses are different. This is because  $G_s$  according to (10) does not abruptly stop transformation, but rather provides an energetic penalty that makes it harder. Further, the loading and unloading portions at the highest stresses do not necessarily reflect the true elastic modulus of the martensite, but rather a combination of elastic and inelastic response. Finally, the apparent moduli at the highest stresses are different in loading and unloading. These modeled behaviors are all consistent with experimental observations (Pelton et al., 2016).

Figs. 6(b,g) show the effect of reducing the power  $m_s$  to 3.0 (the rest of the parameters are as in Table 1). Note that the post-saturation responses are not as smooth and have apparent slopes that are more similar to the pre-saturation responses. Figs. 6(c,h) show the effect of increasing the coefficient  $k_s$  to 0.05 ( $m_s=3$  and rest as in Table 1): we see that the post-saturation slopes become more different from the pre-saturation behaviors. Figs. 6(d,i) show that changing  $f_{\lambda\lambda}$  changes the hardening slope for both pre- and post-saturation transformation, while (e,j) show that  $n_{\lambda\lambda}$  changes the shape of these hardening regions. While we do not explicitly show it here, we will later document that changing  $m_s$  and  $k_s$  uniquely changes the hardening for different loading modes while  $f_{\lambda\lambda}$  and  $n_{\lambda\lambda}$  do so equally in all loading modes. Again, these differences arise from  $f_{\lambda\lambda}$  and  $n_{\lambda\lambda}$  acting on  $\lambda$ , which is always evolving between 0 and 1 during transformation regardless of loading mode or direction, while  $m_s$  and  $k_s$  are only active when  $\lambda \epsilon_m$  takes on values outside of the set  $\mathcal{G}_s$ , which may be both asymmetric and anisotropic.

## 5.2. Uniaxial superelasticity

We now fit and compare our model to experimental observations. Again, one of the key distinguishing features of the model from previous literature is the distinction between initiation and saturation. An important consequence of this is the fact that the transformation does not complete in superelastic loading. So we begin our comparison with experiments by studying this aspect relative to the recent in-situ neutron diffraction experiments of Stebner et al. (2015) where they subjected polycrystalline Nitinol specimens to uniaxial tension and compression while simultaneously measuring the volume fraction of martensite. Their forward loading results are shown in Fig. 7 as the dotted blue lines. In particular, they loaded the specimens well past-saturation. In fact, they had a large amount of plastic deformation. We do not show plasticity data for this comparison since this model does not include plasticity. Still, they found that the martensite phase fraction failed to reach  $\lambda = 1$ . In contrast, they confirmed that the transformation went to completion on cooling (within the threshold of the measurement technique, which is  $\sim 3\%$  martensite).

Fig. 7 also shows the results of our model parameterized according to Table 1. In comparing phase fraction evolutions of our simulation results (solid black lines) with experiments in Fig. 7b, it is apparent that our model captures the very important, aforementioned aspect of the physics of transformation – transformation is not complete at the end of the superelastic plateaus, and some of the material fails to transform, even in further superelastic loading. Notice that the model is also able to capture the superelastic stress–strain behavior on loading (Fig. 7a): the initial elastic response, the onset of transformation, the apparent yield, the saturation as well as the tension–compression asymmetry in critical stress, transformation strain and the slope of the saturation curves. We do not show the experimental unloading curves since they were loaded to high enough stresses to cause plastic deformation and our model does not include plastic deformation. We will



**Fig. 12.** Comparisons of model simulations (black solid lines) to the non-proportional tension-then-shear experimental data of McNaney et al. (colored broken lines). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

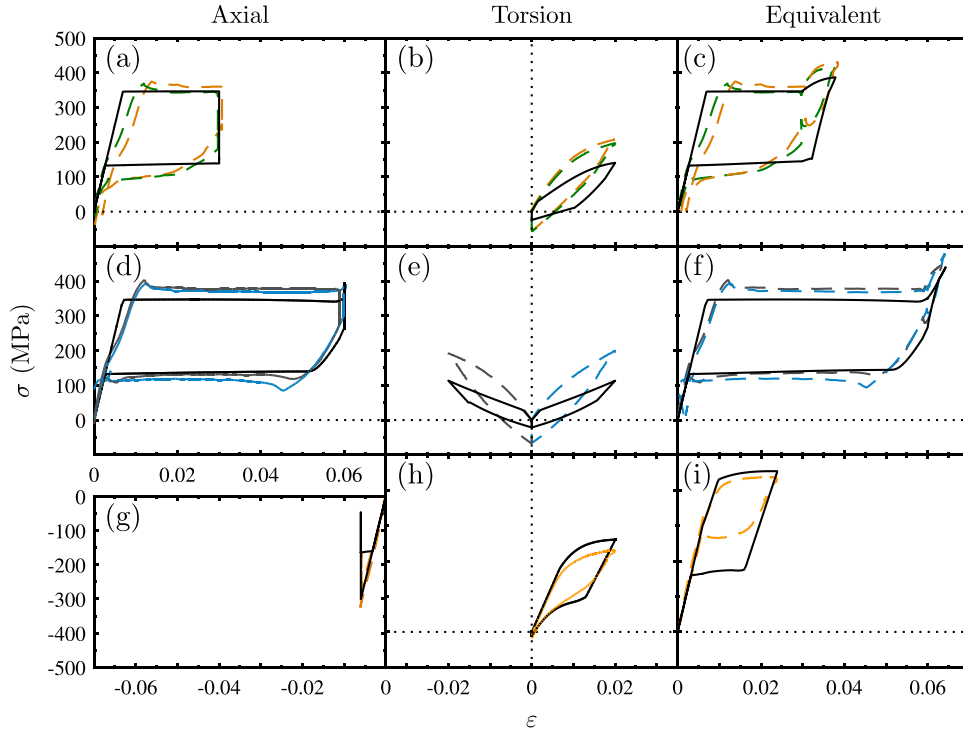
conduct a complete comparison in our future work that includes plastic deformation.

### 5.3. Multiaxial superelasticity of anisotropic tubes

A second consequence of our distinct treatment of initiation and saturation is the natural emergence of martensite reorientation. This reorientation depends on loading mode and in turn can give rise to complex stress–strain curves in multiaxial loading. We examine the ability of the model in these aspects by comparing it to the axial-torsion experiments of McNaney et al. (2003).<sup>5</sup> They subjected Nitinol tubes to axial tension and compression, to torsion and to combined axial-torsion loading both proportionally and non-proportionally in displacement control. They assumed that the state of stress and strain was uniform through the thickness and provided axial stress vs. strain ( $\sigma_{zz}$  vs.  $\epsilon_{zz}$ ), shear stress vs. strain ( $\sigma_{\theta z}$  v.  $\epsilon_{\theta z}$ ) and also (Mises) equivalent stress vs. equivalent strain data. We follow their convention in defining equivalent stress and strain:  $\sigma_{eq} = \sqrt{\sigma_{axial}^2 + 3\sigma_{torsion}^2}$ ,  $\epsilon_{eq} = \sqrt{\epsilon_{axial}^2 + \frac{4}{3}\epsilon_{torsion}^2}$ .

We fit the initiation and saturation surface functions to the results of a few proportional tests. In doing so, we allow the transformation responses of the material to be orthotropic in a cylindrical coordinate system (i.e., unique axial, hoop, and

<sup>5</sup> We thank J.M. McNaney for sharing additional raw data.



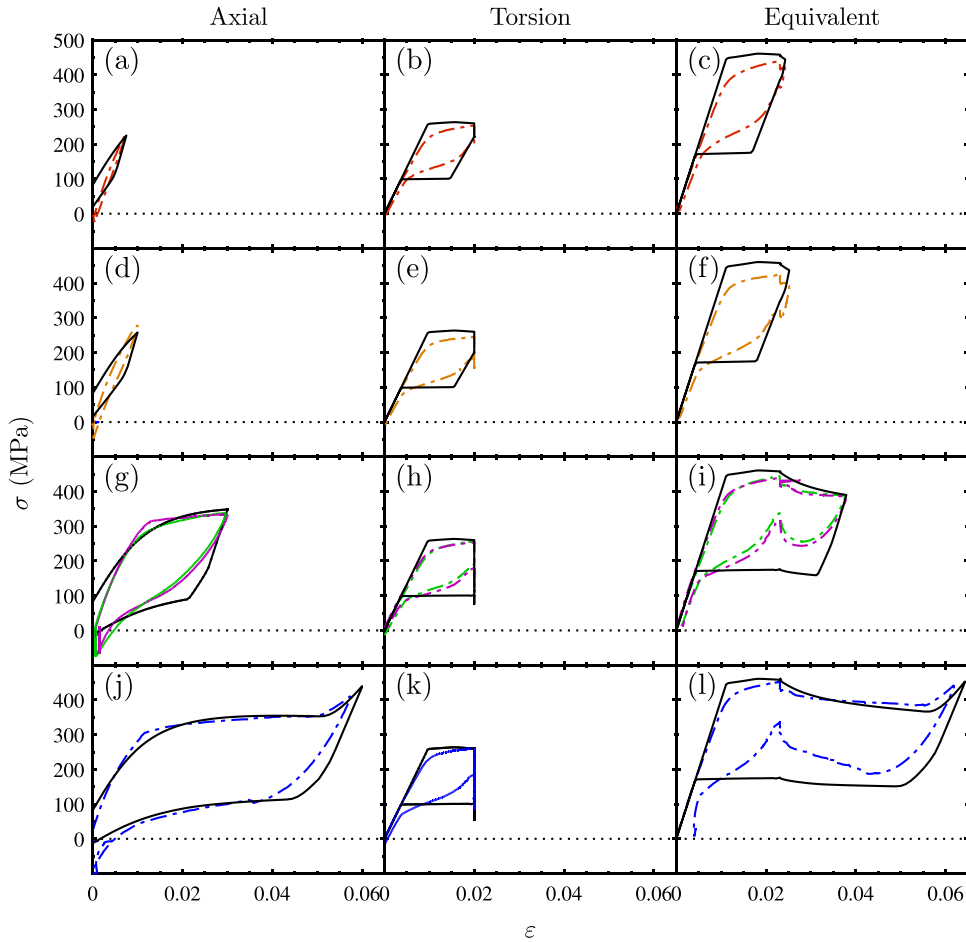
**Fig. 13.** Comparisons of model simulations (black solid lines) to (a)–(f) further non-proportional tension-then-shear, as well as (g)–(i) compression-then-shear experimental data of McNaney et al. (colored broken lines). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

radial responses), as is expected of cold-drawn tubes. We also fit the hardening and kinetic coefficients to results of proportional tests, except for  $d_m^c$ . We fit this parameter according to results of non-proportional loading – namely the emergence of secondary hysteresis loops. The resulting parameters are given in Table 2. Two-dimensional cross-sections of the resulting initiation and saturation surfaces, as well as the measured data points used for fitting them are shown in Fig. 8.

Fig. 9(a) shows the loading direction in the various proportional tests conducted by MacNaney et al. The corresponding observations are shown in Fig. 10 along with the results of the simulations using the parameters in Table 2. Fig. 9(b) shows the resulting nominal transformation strain according to the model. Note that the nominal transformation strain is not necessarily proportional to the total applied strain. Further, in every test except the purely axial strain test, there is significant amounts of martensite reorientation as anticipated by the analysis in Section 3.2.2. Turning now to the stress–strain curves in Fig. 10, we see that the model captures the distinguishing, complex features of the experimental data quite well. Specifically, note that it captures the different plateau stresses in the different loading directions both in loading and unloading. Second, note that in (g,h,i), the shear stress vs. shear strain curve seems to show a negative slope, and this is captured by the model. Further, the model also shows that this is a result of martensite reorientation (red curve in Fig. 9). (j, k,l) and (m,n,o) show the complementary situation of apparent softening in the uniaxial stress–strain curve (green and cyan curves in Fig. 9). There are some subtle features that are not fully captured by the model, notably the pronounced increase in shear hardening in Fig. 10 (k, n). Qualitatively, the model still shows hardening, just not to the same extent. The reason for this disagreement is unclear and a careful examination of such tests both using in-situ measurements and full micro-mechanical simulations will be useful in further elucidating the responsible mechanisms. It is also possible that a complete numerical optimization of the model parameterization could arrive at a solution we did not stumble upon in our manual, instinctual efforts.

We now turn to the non-proportional loading following two protocols: tension-then-shear and shear-then-tension as shown in Fig. 11(a). The corresponding evolution of the nominal transformation strain is shown in Fig. 11(b); as in proportional loading, the nominal transformation strain is not proportional to the applied strain and also shows significant martensite reorientation. However, the corresponding volume fractions evolve similarly as a function of equivalent transformation strains (Fig. 11(c)). The mechanics of the nominal transformation strain dominating the path-dependence of the material responses to non-proportional loading result from choosing  $d_m^c \neq 0$  while simultaneously prescribing  $C_i = C_s$ . Should further micromechanical investigations reveal that phase fraction evolution is also path-dependent, unique choices of  $C_i \neq C_s$  may be made to introduce appropriate path-dependent mechanics of martensite phase fraction evolution to the model, as was discussed in Section 3.2.2.

Figs. 12 and 13 show the experimental stress–strain behavior for axial-then-shear loadings (all tension-then-shear tests



**Fig. 14.** Comparisons of model simulations (black solid lines) to the non-proportional shear-then-tension experimental data of McNaney et al. (colored broken lines). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

except one compression-then-shear test in Figs. 13g,h,i), and compare the data with the simulations using our model. Fig. 14 show the complementary comparisons for the shear-then-tension tests. Remarkable agreement is observed, especially considering that the model was manually fit to the proportional loading tests, except for the parameter  $d_m^c$ , which was the only parameter further tuned to these data. Specifically, the model is able to capture the diverse stress plateaus. Further, the model accurately captures the hardening when one switches from tension to shear (Figs. 12 and 13) and the softening when one switches from shear to tension (Fig. 14). As before, the model misses some subtle details. One known reason is the presence of the R-phase transformation in these experiments, which our model neglects (see Fig. 6 of Stebner et al., 2015, which shows in-situ diffraction data on the tubes used by McNaney et al.). Other possible reasons are also worth investigating further through detailed in-situ experiments and higher fidelity micromechanical simulations, as well as the development of a numerical optimization scheme for parameterizing this model.

## 6. Discussion

We have presented a macroscopic constitutive model of shape-memory alloys that is capable of describing both thermal and stress-induced transformation; the initiation, reorientation and saturation of martensite under various conditions; the difference between various loading modes including tension–compression–shear asymmetry and material anisotropy. While various aspects of the model, such as the initiation and saturation surfaces, have an underlying micromechanical basis, they are presented in a form that can be empirically fit from an experiment, and also easily and efficiently implemented into engineering design tools, such as commercial finite element software.

We conclude with a discussion of a few of the aspects that have been omitted and possible extensions of the model.

- **Plasticity:** The commonly used shape-memory alloy Nitinol is quite ductile due to both slip and deformation twinning. It is typically precipitate and/or work hardened so that the yield stress is sufficiently high that it prevents homogeneous slip

from overshadowing the shape-memory effect and superelasticity. Still, it is common to see some heterogeneous plasticity accompany transformation, especially in the first few cycles. We can extend the model presented here to incorporate plasticity following the approach in [Sadjadpour and Bhattacharya \(2007b\)](#). Briefly, the plastic strain is introduced as an internal variable that is additive to the elastic and transformation strain, and the yield behavior is incorporated as an evolution law for this internal variable. However, such a formulation does not account for interactions between plasticity and transformation. The interaction between these phenomena is incompletely understood. Still, in a companion article, we will present a new formulation that accounts for coupling between transformation and plasticity based on our best ideas ([Stebner and Bhattacharya, 2016](#)).

- *Finite element implementation:* The model and the incremental variational formulation given in Section 4.1 are amenable to numerical implementation for large scale simulation. [Kelly \(2008\)](#) incorporated an earlier version of this model as a user-defined material (UMAT) in the commercial finite-element analysis package ABAQUS, and used it to study various examples. Such an implementation may be able to resolve disagreements in our comparison to the McNaney et al. data with regard to the Poynting effect ([Poynting, 1909](#)), i.e., the accumulation of axial stress during a strain controlled pure twist of a tube or vice-versa, which is much more pronounced in the experiments than the simulations. In our material point code, we are using strain-control boundary conditions on the loading axes and completely ignoring the structural mechanics of tubes, grips, and the tube specimen geometry, where the real experiment uses load/displacement conditions. A structural implementation could also simulate other phenomena known of SMA specimens, such as strain localization.

## Acknowledgment

This work draws from the doctoral thesis of AK at the California Institute of Technology. AK and KB gratefully acknowledge the financial support of the US Army Research Office (W911NF-07-1-0410) and US Department of Energy (DE-FC52-08NA28613). AS acknowledges support from US Department of Energy - National Nuclear Security Administration (DE-NA0001805) that supported his post-doctoral studies at California Institute of Technology, where he began his efforts on this project, and thanks the Colorado School of Mines for support that enabled its completion.

## Appendix A. Derivatives of principal values and directions

We use eigenvalue based functions for initiation and saturation ([7](#)), and need to differentiate these functions to implement the model (e.g., [\(84\)](#) and [\(83\)](#)). Hence, we recall explicit formulae for differentiation of eigenvalues ( $\mu$ ) and eigenvectors ( $x$ ) with respect to a symmetric,  $3 \times 3$  matrix  $A$ .

We seek:

$$\frac{d\mu}{dA}, \quad \frac{dx}{dA}, \quad \text{where } Ax = \mu x, \quad |x| = 1. \quad (91)$$

Perturb the above equation:

$$(A + \varepsilon B)(x + \varepsilon y) = (\mu + \varepsilon \lambda)(x + \varepsilon y). \quad (92)$$

Terms of  $O(1)$  recover the original equation. So looking at  $O(\varepsilon)$ ,

$$Ay + Bx = \mu y + \lambda x. \quad (93)$$

Take the dot product with  $x$  and note that  $(Ay) \cdot x = y \cdot A^T x = y \cdot Ax = \mu(y \cdot x)$  so that

$$\frac{d\mu}{dA}: B = \lambda = x \cdot Bx \quad \text{or} \quad \frac{d\mu}{dA} = x \otimes x. \quad (94)$$

Further, from [\(93\)](#),

$$(A - \mu I)y = -Bx + x(x \cdot Bx), \quad (95)$$

or

$$(A - \mu I)y = -PBx \quad \text{where } P = I - x \otimes x \quad (96)$$

is the projection to the linear space perpendicular to  $x$ . We solve this equation above for  $y = \frac{dx}{dA} \cdot B$ . This problem is well-posed when  $\mu$  is a distinct eigenvalue since  $PBx$  belongs to the linear space perpendicular to  $x$ . Formally,

$$\frac{dx}{dA}: B = -(A - \mu I)^{-1}PBx \quad \text{or} \quad \frac{dx}{dA} = -(A - \mu I)^{-1}P \otimes x. \quad (97)$$

For the second derivatives, we may proceed by differentiating the terms above, or perturbing the equation to second

order:

$$(A + \varepsilon B)(x + \varepsilon y + \varepsilon^2 z) = (\mu + \varepsilon \lambda + \varepsilon^2 \kappa)(x + \varepsilon y + \varepsilon^2 z). \quad (98)$$

The  $O(1)$  and  $O(\varepsilon)$  are satisfied identically and therefore we focus on the  $O(\varepsilon^2)$  term:

$$Az + By = \mu z + \lambda y + \kappa x. \quad (99)$$

Take the dot product with  $x$  and note as before that  $(Az) \cdot x = \mu(z \cdot x)$ . We obtain the second variation

$$B: \frac{d^2 \mu}{dAdA}; B = \kappa = x \cdot By - \lambda y \cdot x = y \cdot PBx. \quad (100)$$

## References

- Abeyaratne, Rohan, Chu, C., James, Richard D., 1996. Kinetics of materials with wiggly energies: theory and application to the evolution of twinning microstructures in a Cu–Al–Ni shape memory alloy. *Philos. Mag. A* 73 (2), 457–497.
- Auricchio, F., Petrini, L., 2004. A three-dimensional model describing stress-temperature induced solid phase transformations: solution algorithm and boundary value problems. *Int. J. Numer. Methods Eng.* 61 (December (6)), 807–836.
- Auricchio, F., Taylor, R.L., Lubliner, J., 1997. Shape-memory alloys: macromodelling and numerical simulations of the superelastic behavior. *Comput. Methods Appl. Mech.* 146 (3–4), 281–312.
- Bhattacharya, K., Kohn, R.V., 1997. Energy minimization and the recoverable strains of polycrystalline shape-memory alloys. *Arch. Ration. Mech. Anal.* 139, 99–180.
- Bhattacharya, K., Schlömerkemper, A., 2004. Transformation yield surface of shape memory alloys. *J. Phys. IV* 115 (January), 155–162.
- Bhattacharya, K., Schlömerkemper, A., 2010. Stress-induced phase transformations in shape-memory polycrystals. *Arch. Ration. Mech. Anal.* 196 (January (3)), 715–751.
- Bhattacharya, K., Suquet, P.M., 2005. Model problem concerning recoverable strains of shape-memory polycrystals. *Proc. R. Soc. Lond. A* 461, 2797–2816.
- Brinson, L.C., Schmidt, I., Lammering, R., 2004. Stress-induced transformation behavior of a polycrystalline niti shape memory alloy: micro and macromechanical investigations via in situ optimal microscopy. *J. Mech. Phys. Solids* 52, 1549–1571.
- Burkart, M.W., Read, T.A., 1953. Diffusionless phase change in the indium-thallium system. *Trans. Am. Inst. Mining Metall. Eng.* 197, 1254–1516.
- Cazacu, Oana, Plunkett, Brian, Barlat, Frédéric, 2006. Orthotropic yield criterion for hexagonal closed packed metals. *Int. J. Plast.* 22 (7), 1171–1194.
- Chemisky, Y., Duval, A., Patoor, E., Ben Zineb, T., 2011. Constitutive model for shape memory alloys including phase transformation, martensitic reorientation and twins accommodation. *Mech. Mater.* 43 (December (7)), 361–376.
- Coleman, B.D., Noll, W., 1963. The thermodynamics of elastic materials with heat conduction and viscosity. *Arch. Ration. Mech. Anal.* 13 (3), 167–178.
- Daly, S., Bhattacharya, K., Ravichandran, G., 2007. Stress-induced martensitic phase transformation in thin sheets of nitinol. *Acta Mater.* 55, 3593–3600.
- Duerig, T.W., Bhattacharya, K., 2015. The influence of the R-phase on the superelastic behavior of Niti. *Shape Mem. Superelast.* 1, 153–161.
- Ekeland, I., Témam, R., 1999. *Convex Analysis and Variational Problems*. SIAM, Philadelphia.
- James, R.D., 1986. Displacive phase-transformations in solids. *J. Mech. Phys. Solids* 34 (January (4)), 359–394.
- Kelly, A., 2008. *A Constitutive Relation for Shape-Memory Alloys* (Thesis). California Institute of Technology.
- Lexcellent, C., Vivet, A., Bouvet, C., Calloch, S., Blanc, P., 2002. Experimental and numerical determinations of the initial surface of phase transformation under biaxial loading in some polycrystalline shape-memory alloys. *J. Mech. Phys. Solids* 50 (December (12)), 2717–2735.
- McNaney, J.M., Imbeni, V., Jung, Y., Papadopoulos, P., Ritchie, R.O., 2003. An experimental study of the superelastic effect in a shape-memory nitinol alloy under biaxial loading. *Mech. Mater.* 35, 969–986.
- Mielke, A., 2005. *Evolution of Rate-Independent Systems*, vol. 2. .
- Miyazaki, S., Kimura, S., Otsuka, K., Suzuki, Y., 1984. The habit plane and transformation strains associated with the martensitic transformation in Ti–Ni single crystals. *Scr. Metall. Mater.* 18 (January (9)), 883–888.
- Otsuka, K., Wayman, C.M., 1998. *Shape Memory Materials*. Cambridge University Press, Cambridge.
- Panico, M., Brinson, L.C., 2007. A three-dimensional phenomenological model for martensite reorientation in shape memory alloys. *J. Mech. Phys. Solids* 55 (January (11)), 2491–2511.
- Paranjape, Harshad, Anderson, Peter M., 2014. Texture and grain neighborhood effects on Ni–Ti shape memory alloy performance. *Model. Simul. Mater. Sci. Eng.* 22 (7), 075002.
- Patel, J.R., Cohen, M., 1953. Criterion for the action of applied stress in the martensitic transformation. *Acta Metall.* 1 (5), 531–538.
- Pelton, A.R., Clausen, B., Stebner, A.P., 2015. In-situ neutron diffraction studies of increasing tension pre-strain amplitudes of superelastic nitinol. *Shape Mem. and Superelast.* 1, 375–386.
- Poynting, J.H., 1909. On pressure perpendicular to the shear-planes in finite pure shears, and on the lengthening of loaded wires when twisted. *Proc. R. Soc. Lond. A* 82, 546–559.
- Poynting, J.H., 1909. On pressure perpendicular to the shear planes in finite pure shears, and on the lengthening of loaded wires when twisted. *Proc. R. Soc. Lond. Ser. A, Contain. Pap. Math. Phys. Character* 82 (557), 546–559.
- Qidwai, M.A., Lagoudas, D.C., 2000. Numerical implementation of a shape memory alloy thermomechanical constitutive model using return mapping algorithms. *Int. J. Numer. Methods Eng.* 47 (January (6)), 1123–1168.
- Richards, A.W., Lebensohn, R.A., Bhattacharya, K., 2013. Interplay of martensitic phase transformation and plastic slip in polycrystals. *Acta Mater.* 61 (12), 4384–4397.
- Sadjadpour, A., Bhattacharya, K., 2007a. A micromechanics-inspired constitutive model for shape-memory alloys. *Smart Mater. Struct.* 16, 1751.
- Sadjadpour, A., Bhattacharya, K., 2007b. A micromechanics inspired constitutive model for shape-memory alloys: the one-dimensional case. *Smart Mater. Struct.* 16, S51.
- Shield, T.W., 1995. Orientation dependence of the pseudoelastic behavior of single-crystals of Cu–Al–Ni in tension. *J. Mech. Phys. Solids* 43, 869–895.
- Shu, Y.C., Bhattacharya, K., 1998. The influence of texture on the shape-memory effect in polycrystals. *Acta Mater.* 46, 5457–5473.
- Sitter, P., Paidar, V., Letter, L., Seiner, H., 2010. In: *Proceedings of the 8th European Symposium on Martensitic Transformations*. EDP Sciences.
- Sittner, P., Novak, V., 2000. Anisotropy of martensitic transformations in modeling of shape memory alloy polycrystals. *Int. J. Plast.* 16 (January (10–11)), 1243–1268.
- Stebner, A.P., Bhattacharya, K., 2016. A micromechanics inspired constitutive model for shape-memory alloys that accounts for coupling of phase transformation and plasticity. *J. Mech. Phys. Solids*, Submitted for publication.
- Stebner, A.P., Paranjape, H.M., Clausen, B., Brinson, L.C., Pelton, A.R., 2015. In-situ neutron diffraction studies of large monotonic deformations of superelastic nitinol. *Shape Mem. Superelast.* 1, 252–267.
- Stebner, A.P., Vogel, S.C., Noebe, R.D., Sisneros, T.A., Clausen, B., Brown, D.W., Garg, A., Brinson, L.C., 2013. Micromechanical quantification of elastic, twinning, and slip strain partitioning exhibited by polycrystalline, monoclinic nickel–titanium during large uniaxial deformations measured via in-situ neutron diffraction. *J. Mech. Phys. Solids* 61 (11), 2302–2330.
- Thamburaja, P., Anand, L., 2002. Superelastic behavior in tension–torsion of an initially-textured Ti–Ni shape-memory alloy. *Int. J. Plast.* 18 (11), 1607–1617.
- Yang, Q., Stainier, L., Ortiz, M., 2006. A variational formulation of the coupled thermo-mechanical boundary-value problem for general dissipative solids. *J. Mech. Phys. Solids* 54 (January (2)), 401–424.
- Zaki, W., Moumni, Z., 2007. A 3d model of the cyclic thermomechanical behavior of shape memory alloys. *J. Mech. Phys. Solids* 55 (11), 2427–2454.